TERRESTRIAL ARTHROPOD BIODIVERSITY
ON THE KENAI NATIONAL WILDLIFE REFUGE, ALASKA

By

Matthew L. Bowser

RECOMMENDED:

_____

_____

_____
Advisory Committee Chair

_____
Chair, Department of Biology and Wildlife

APPROVED:

_____
Interim Dean,
College of Natural Science and Mathematics

_____
Dean of the Graduate School

_____
Date

TERRESTRIAL ARTHROPOD BIODIVERSITY ON THE KENAI NATIONAL WILDLIFE REFUGE, ALASKA

A

THESIS

Presented to the Faculty
of the University of Alaska Fairbanks
in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

By

Matthew L. Bowser

Fairbanks, Alaska

May 2009

Abstract

Elucidating the causes of observed patterns of living diversity remains a central goal of ecology. To understand patterns of terrestrial arthropod diversity on Kenai National Wildlife Refuge (KENWR), arthropods were collected by sweep net on 255 100m^2 plots systematically distributed at 4.8km intervals across KENWR. I calculated three indices of diversity for 90 families conveying information on richness and evenness for each site. Using Bayesian Model Averaging, I found all indices were strongly influenced by site productivity, local climate, time of sampling and plant species richness. Physiographic variables were less important than climate for determining arthropod distributions.

Because many species are expected to alter their distributions in response to accelerated climate change, I assessed the use of occupancy models for monitoring those shifts on KENWR. I compared rotating panel and periodic census sampling designs using Monte-Carlo simulations given a range of occupancy and detectability values. Both designs estimated detectability within single visits and provided reasonable precision and accuracy on occupancy estimates of species with detection probabilities ≥ 0.5, but the rotating panel design was preferred because it yielded information at shorter time intervals. I recommended adding sites sampled in consecutive seasons to better estimate local extinction and colonization rates.

Table of Contents

Page

Signature Page . . . . . **i**

Title Page . . . . . **ii**

Abstract . . . . . **iii**

Table of Contents . . . . . **iv**

List of Figures . . . . . **viii**

List of Tables . . . . . **x**

List of Appendices . . . . . **xii**

Acknowledgments . . . . . **xiii**

1 Arthropod Family Diversity **1**

1.1 Introduction 1

Justifications for Studying Arthropod Diversity on the Kenai National Wildlife Refuge 1

Relevant Patterns of Diversity 3

Relationships of Diversity with Latitude and Elevation: The Importance of Climate 3

Relationships of Diversity with Productivity 6

Congruence of Diversity Patterns of Arthropods and Plants 7

Practical Considerations in Measurement of Arthropod Diversity 8

Taxonomic Resolution 8

Taxonomic Breadth 10

Objectives and Hypotheses 11

Objective 1: Document patterns of terrestrial arthropod diversity. . 11

Objective 2: Explain the causes of the observed pattern of diversity. 11

1.2 Methods 12

Page

Overview . . . 12
Study Area . . . 12
General Description . . . 12
Climate . . . 15
Introduction to LTEMP . . . 16
Sampling Design . . . 16
Vegetation Sampling . . . 17
Plot Weather Data . . . 17
Arthropod Sampling . . . 18
Arthropod Sampling Methods . . . 19
Taxonomic Resolution and Breadth . . . 20
Sorting and Identification . . . 20
GIS-derived Data . . . 20
Elevation Data . . . 20
Productivity Data . . . 22
Climate data . . . 24
Rarefaction . . . 24
Prediction by Random Forest Regression . . . 24
Diversity Indices . . . 25
Analyses of Spatial Autocorrelation . . . 27
Regression Analyses of Diversity Indices . . . 27
Regressions of Observed Temperatures and NDVI . . . 31
1.3 Results . . . 31
Quantity of Material Collected . . . 31
Taxa Collected . . . 32
Diversity . . . 33
Random Forest Regression of Family Density . . . 36
Regression Analyses of Family Density . . . 40
Regression Analyses of Shannon's Information Index . . . 43
Regression Analyses of the Berger-Parker Index . . . 45
Comparison of Diversity Indices . . . 47

Page

Regressions of Observed Temperatures and NDVI 49
Spatial Autocorrelation of Residuals 50
1.4 Discussion 51
Overview of the Fauna Collected 51
Overview of Diversity Patterns 52
Seasonal Phenology, Daily Activity Patterns, and Temperature 52
Precipitation 54
Elevation 54
Productivity 55
Vascular Plant Species Density 56
The Hemlock Forest Habitat Type 56
The Alpine Habitat Type 56
The Halophytic Wetland Habitat Type 57
Spatial Autocorrelation 57
Prediction by Random Forest Regression 58
Synthesis 58
Concerns/Caveats 60

2 Monitoring of Terrestrial Arthropods on the KENWR 63
2.1 Introduction 63
Distribution shifts due to changing climate 63
Documenting species distributions 65
Monitoring species distributions on KENWR's LTEMP 67
Objectives 70
2.2 Methods 70
Evaluating LTEMP monitoring designs 70
Generating simulated data 71
Obtaining naïve estimates of occupancy 73
Fitting occupancy models using program PRESENCE 73
Single-season scenarios 74
Short-term, multi-season scenarios 74

Page

Long-term, multi-season scenarios 75
Dealing with convergence failures 75
Evaluating Arthropod Taxa as Monitoring Candidates. 76
2.3 Results 76
Evaluation of LTEMP monitoring designs 76
Single-season estimates 76
Multi-season estimates: short term performance 84
Multi-season estimates: long term performance 87
Evaluation of arthropod taxa as monitoring candidates. 91
2.4 Discussion 92
Evaluation of LTEMP monitoring designs 92
Software and failure rates 93
Evaluation of arthropod taxa as monitoring metrics 94
Conclusions 99

References **100**

List of Figures

Page

1.1 Generalized climatic model for elevational gradients in species richness of bats from McCain (2007). 6
1.2 Information retained vs. cost of obtaining taxonomic data. 10
1.3 Map of the Kenai National Wildlife Refuge. 14
1.4 Kenai Peninsula land cover classification (O'Brien, 2006). 15
1.5 LTEMP sampling points. 19
1.6 NDVI raster generated from LandSat 7 imagery. 23
1.7 Numbers of specimens collected per plot 32
1.8 Rarefaction curve of families collected. 33
1.9 Histograms of diversity indices. 34
1.10 Spatial autocorrelation of diversity indices: empirical semivariograms. 36
1.11 Arthropod family density predicted by random forest regression. 37
1.12 Standard error of arthropod family density predicted by random forest regression. 38
1.13 Importance of variables included in final random forest regression model. 39
1.14 Predicted family density vs. total annual precipitation at varying values of observed plot temperature. 42
1.15 Predicted family density vs. total annual precipitation at varying values of NDVI. 43
1.16 Predicted values of Shannon's Information Index over the range of values of observed plot temperature and time of day. 45
1.17 Arthropod family density versus total annual precipitation. 49
1.18 Spatial autocorrelation of residuals of diversity models: empirical semivariograms. 51

2.1 Failure rate of Program PRESENCE over a range of values of Ψ and p for the rotating panel design. 77
2.2 Failure rate of Program PRESENCE over a range of values of Ψ and p for the periodic census design. 78
2.3 Bias of naïve estimates of Ψ for the rotating panel design. 79
2.4 Bias of naïve estimates of Ψ for the periodic census design. 79
2.5 Bias of PRESENCE estimates of Ψ for the rotating panel design. 80

Page

2.6 Bias of PRESENCE estimates of Ψ for the periodic census design. 81
2.7 Standard deviation of PRESENCE estimates of Ψ for the rotating panel design. 82
2.8 Standard deviation of PRESENCE estimates of Ψ for the periodic census design. 82
2.9 Standard deviation of naïve estimates of Ψ for the rotating panel design. 83
2.10 Standard deviation of naïve estimates of Ψ for the periodic census design. . . . 83
2.11 Failure rate of program PRESENCE over 4 years for the rotating panel design. . 84
2.12 Mean bias of estimates of the rate of local extinction for the rotating panel design over four years. 85
2.13 Mean bias of estimates of the local rates of colonization for the rotating panel design over four years. 86
2.14 Standard deviation of the local rate of extinction for the rotating panel design over four years. 86
2.15 Standard deviation of estimates of the local rates of colonization for the rotating panel design over four years. 87
2.16 Failure rate of program PRESENCE over 20 years of the rotating panel design. . 88
2.17 Estimates of the local rate of extinction over 20 years for the rotating panel design. 89
2.18 Estimates of the local rate of colonization over 20 years for the rotating panel design. 89
2.19 Estimates of the local rate of colonization over 20 years for the periodic census design. 90
2.20 Estimates of the local rate of colonization over 20 years for the periodic census design. 91

List of Tables

Page

1.1 Sky condition codes and definitions. 18
1.2 Arthropod families included in analyses. 21
1.3 Covariates included in random forest regressions. 25
1.4 Terms included in BMA Analyses. 29
1.5 Correlation matrix of diversity indices and plot sums. 35
1.6 Spatial autocorrelation of diversity indices: results of inverse-distance-weighted Moran's I tests. 35
1.7 Predicted arthropod family density by landcover type. 40
1.8 Best five negative binomial regression models of family density selected by Bayesian Model Averaging. 41
1.9 Results of negative binomial regression model selected by BMA analysis. 42
1.10 Best five Gaussian ordinary least squares regression models of Shannon's Information Index selected by Bayesian Model Averaging. 44
1.11 Results of Gaussian oridanary least squares regression model selected by BMA analysis. 45
1.12 Best five gamma regression models of the Berger-Parker Index selected by Bayesian Model Averaging. 46
1.13 Results of gamma regression model selected by BMA analysis. 47
1.14 Comparison of variables in regression coefficients most often chosen by BMA analyses. 48
1.15 Results of regression of observed temperatures. 50
1.16 Spatial autocorrelation of residuals of diversity models: results of inverse-distance-weighted Moran's I tests. 51

2.1 Tabular representation of proposed LTEMP sampling designs over 20 years . . 69
A.1 Full results of BMA analyses: family richness. 114
A.2 Full results of BMA analyses: Shannon's Information Index. 116
A.3 Full results of BMA analyses: the Berger-Parker Index. 118
B.1 Frequencies and quantities of arthropod families collected. 120

Page

C.1 Failure rate of the rotating panel design for a single season. 124
C.2 Failure rate of the periodic census design for a single season. 124
C.3 Bias of estimates of Ψ for the rotating panel design for a single season, naïve estimates. 125
C.4 Bias of estimates of Ψ for the periodic census design for a single season, naïve estimates. 125
C.5 Bias of estimates of Ψ for the rotating panel design for a single season, PRESENCE estimates. 126
C.6 Bias of estimates of Ψ for the periodic census design for a single season, PRESENCE estimates. 126
C.7 Standard deviation of estimates of Ψ for the rotating panel design for a single season, naïve estimates. 127
C.8 Standard deviation of estimates of Ψ for the periodic census design for a single season, naïve estimates. 127
C.9 Standard deviation of estimates of Ψ for the rotating panel design for a single season, PRESENCE estimates. 128
C.10 Standard deviation of estimates of Ψ for the periodic census design for a single season, PRESENCE estimates. 128
C.11 Failure rate of the rotating panel design over four years. 129
C.12 Bias of estimates of ε for the rotating panel design over four years, PRESENCE estimates. 130
C.13 Standard Deviation of estimates of ε for the rotating panel design over four years, PRESENCE estimates. 130
C.14 Bias of estimates of γ for the rotating panel design over four years, PRESENCE estimates. 131
C.15 Standard deviation of estimates of γ for the rotating panel design over four years, PRESENCE estimates. 131

List of Appendices

Page

A Full results of Bayesian Model Averaging analyses 113

B Summary of arthropods collected 120

C Tabular results of simulations 124

D R code used for running occupancy simulations using program PRESENCE . . . 132

Acknowledgments

I wish to thank Pat Doak (University of Alaska Fairbanks), Diane Wagner (University of Alaska Fairbanks), and John Morton (US Fish & Wildlife Service) for their help as my graduate advisers. The USDA Forest Inventory and Analysis program graciously allowed us to work on their grid and shared their data. The USFWS Kenai National Wildlife Refuge provided most of the funding for this project. The University of Alaska Fairbanks provided support through teaching assistantships. Marathon Oil donated additional funding. My wife deserves much thanks for her patience and encouragement throughout this project. I owe special thanks to Dominique Collet for sharing his laboratory and for diverse kinds of assistance. Charlotte Hockin performed a large part of the rough sorting of specimens. Todd Eskelin (US Fish & Wildlife Service) helped in various ways. My parents provided encouragement and support. Many (listed below) have helped with identifications.

Armitage, Brian (Trichoptera). Midwest Biodiversity Institute, Ohio State University, Columbus, Ohio.

Ball, George E. (Carabidae). Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada.

Barber, Kevin N. (Anthomyzidae). Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada Sault Ste. Marie, Ontario, Canada.

Bernard, Ernest C. (Hypogastruridae, Isotomidae) Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee.

Buffington, Matthew. (Figitidae). Systematic Entomology Lab, USDA Agricultural Research Service Washington, D.C.

Clausen, Philip J. (Ephydridae). Department of Entomology, University of Minnesota St. Paul, Minnesota

Cokendolpher, James C. (Opiliones). Invertebrate Collection, Natural Science Research Laboratory, Museum of Texas Tech University Lubbock, Texas.

Collet, Dominique. (Chalcidoidea, many other groups). Sterling, Alaska

Douglas, Hume. (Elateridae). Entomology, Ontario Plant Laboratories, C.F.I.A., Government of Canada. Ottawa, Ontario, Canada.

Eskelin, Todd. (Odonata). Kenai National Wildlife Refuge, U.S. Fish & Wildlife Service. Soldotna, Alaska

Ferris, Clifford. (Lepidoptera). Bioengineering Program, University of Wyoming. University Station, Laramie, Wyoming.

Fielding, Dennis (Acrididae) USDA-ARS Subarctic Agricultural Research Unit, University of Alaska Fairbanks, Fairbanks, Alaska.

Foote, Benjamin. (Sciomyzidae, Ephydridae, Heleomyzidae). Department of Biological Sciences, Kent State University. Kent, Ohio.

Foottit, Robert. (Psyllidae). Agriculture and Agri-Food Canada. Ottawa, Ontario, Canada

Francoeur, André. (Formicidae). Université du Québec à Chicoutimi, Centre de données sur la biodiversité du Québec. Saguenay, Québec, Canada.

Gordon, Robert. (Scarabaeidae). Northern Plains Entomology. Willow City, North Dakota.

Hagerty, Aaron M. (Coccinellidae). USDA Agricultural Research Service. Fairbanks, Alaska.

Hamilton, K. G. Andy. (Achilidae, Delphacidae). Agriculture & Agri-Food Canada, Central Experimental Farm Ottawa, Ontario, Canada.

Hebert, Paul. (Lepidoptera). Department of Intergrative Biology, University of Guelph. Guelph, Ontario, Canada.

Holmberg, Robert. (Opiliones). Centre for Science, Athabasca University. Athabasca, Alberta, Canada.

Johnson, Bing. (Coniopterygidae). Division of Entomology, Department PSES, University of Idaho. Moscow, Idaho.

Leschen, Richard. (Cryptophagidae). New Zealand Arthropod Collection. Mt. Albert, Aukland, New Zealand.

Mockford, Edward L. (Psocoptera). Department of Biological Sciences, Illinois State University. Normal, Illinois.

Olmi, Massimo. (Dryinidae). Department of Plant Protection, University of Tuscia. Viterbo, Italy.

Ramsdale, Alistair. (Cantharidae, Lampyridae). Montana Entomology Collection, Montana State University. Bozeman, Montana.

Schiff, Nathan M. (Siricidae). Center for Bottomland Hardwoods Research, USDA Forest Service. Stoneville, Mississippi.

Sharkey, Michael. (Braconidae). Department of Entomology, University of Kentucky. Lexington, Kentucky.

Skartveit, John. (Bibionidae). Department of Biology, Realfagbygget. Bergen, Norway.

Skevington, Jeff. (Pipunculidae). Agriculture and Agri-Food Canada. Ottawa, Ontario, Canada.

Smith, David. (Symphyta). Systematic Entomology Laboratory, USDA Agricultural Research Service. Washington, D.C.

Snider, Richard. (Sminthuridae). Department of Zoology, Michigan State University. East Lansing, Michigan.

Steck, Gary. (Tephritidae). Division of Plant Industry, Florida Department of Agriculture & Consumer Services. Gainesville, Florida.

Sturm, Helmut. (Archaeognatha). Universität Hildesheim. Hildesheim, Germany.

Woodley, Norman. (Heleomyzidae). Communications & Taxonomic Services Unit, USDA-ARS-Systematic Entomology Laboratory. Beltsville, Maryland.

Wu, Andrew. (Tetragnathidae). Kenai National Wildlife Refuge, U.S. Fish & Wildlife Service. Soldotna, Alaska.

Zapparoli, Marzio (Lithobiomorpha) Dipartimento di Protezione delle Piante, Viterbo, Italy.

Chapter 1

Arthropod Family Diversity

1.1 Introduction

Justifications for Studying Arthropod Diversity on the Kenai National Wildlife Refuge

Explaining observed patterns of living diversity remains a central goal of ecology despite over a century of study. Since latitudinal diversity gradients were described in the nineteenth century (Wallace, 1853, 1878; Bates, 1862), numerous studies have documented relationships between diversity and latitude, elevation, productivity, and other factors (see Fisher, 1960; Pianka, 1966; Currie, 1991; Rosenzweig, 1995; Lomolino, 2001; Mittelbach et al., 2001; Whittaker et al., 2001; Hillebrand, 2004; McCain, 2005; Mittelbach et al., 2007; Grytnes and McCain, 2007, for reviews and meta-analyses), but satisfactory explanations of the observed patterns remain elusive (Whittaker et al., 2001). Such straightforward questions as "How is diversity related to productivity?" and "Why does the relationship of diversity to elevation vary?" are still contentious, with multiple competing hypotheses. Accordingly, Whittaker et al. (2001) emphasized the need for a synthesis of hypotheses to explain observed patterns of diversity.

Besides explaining observed patterns, the study of diversity has important applications in conservation strategy and for assessment of ecosystem health. Fleishman et al. (2000) argued that documenting patterns of diversity and elucidating the causes of the observed patterns were essential for conservation planning (but see Groves, 2003, for a different opinion). Since ecosystem stability (McCann, 2000), ecosystem function (Hooper et al., 2005) and resilience (Chapin et al., 2000) are generally correlated with diversity, species richness is itself considered to be a good indicator of the health of a system (Magurran, 1988). Under

this premise, taxon richness has been used as an indicator in environmental assessment studies (e.g., Bechtel and Copeland, 1970; Egloff and Brakel, 1973; Wu, 1982; Roth et al., 1994; Karr and Kimberling, 2003).

The simple fact that arthropods comprise the vast majority of diversity in terrestrial systems (Finnamore, 1996) begs their use in studies of diversity. On land, not only do arthropods generally exceed vertebrates in terms of diversity, ubiquity, and biomass, they are also usually more important in terms of ecosystem function (Wilson, 1987). Insects are also important in terms of ecological services rendered (Losey and Vaughan, 2006). Terrestrial arthropods act as decomposers of litter, wood, carcasses, and scat; herbivores of vascular plants, non-vascular plants, and lichens; fungivores; pollinators; dispersers of seeds and other propagules; disease vectors; predators; and parasites. The variety of ecological roles played by arthropods and their ecological importance make terrestrial arthropods particularly meaningful candidates for the study of biodiversity.

In addition to scientific merit and usefulness of arthropod diversity studies, land managers may be required by law to gain an understanding of invertebrate diversity. For example, National Wildlife Refuges in Alaska were mandated in the Alaska National Interest Lands Conservation Act (ANILCA) of 1980, "to conserve fish and wildlife populations in their natural diversity," where "fish and wildlife" was defined as "any member of the animal kingdom including without limitation any mammal, fish, bird, amphibian, reptile, mollusk, crustacean, arthropod, or other invertebrate." ANILCA specifically required the conservation of arthropod diversity on Alaska refuges, yet no Alaska refuge has conducted a thorough inventory of arthropods, much less assessed their conservation statuses. In Alaska, records for most arthropod taxa are incomplete (Danks et al., 1997) so that arthropods are less well known than any other group of land animals (Alaska Department of Fish and Game, 2006).

Box 1. Definitions of Diversity and Spatial Concepts
Whittaker et al. (2001) proposed a unifying nomenclature for diversity and spatial scale concepts, which I will mostly follow. **Species richness**: the number of species present at a given place or region. **Species density**: species richness standardized in some way, such as by area or sampling method. **Evenness**: a measure of how uniformly abundance is distributed among taxa. **Dominance**: the extent to which an assemblage is represented mostly by individuals of a few taxa (the inverse of evenness). Concepts of richness, density, evenness, and dominance can also apply to taxonomic resolutions other than species, such as genus or family richness. **Diversity**: difficult to define (Magurran, 1988), referring to either richness, evenness, or a composite measure of both of these concepts. In this paper, diversity will be used as an inclusive term including both richness (or density) and evenness.
Spatial scale: the size of the sampling unit. **Geographic extent**: the area of inference or its size.

Relevant Patterns of Diversity

Relationships of Diversity with Latitude and Elevation: The Importance of Climate

Latitudinal Gradients One of the most well-known patterns of the diversity of life is the general trend of decreasing diversity progressing from the tropics toward either of the poles, which has been known since the 19$^{\text{th}}$ century (Wallace, 1878). Most groups of plants and animals follow this pattern, both in the present and in the fossil record (Rosenzweig, 1995). Hillebrand (2004), in a meta-analysis of nearly 600 diversity studies, confirmed the generality of this trend. Among terrestrial arthropods, litter mites (Stanton, 1979), dragonflies (Tillyard, 1917; Kennedy, 1928), termites (Collins, 1989), some wood-boring beetles (Beaver, 1979) butterflies (Hovanitz, 1958; Slansky, 1973; Scriber, 1973, 1984), Sphingid moths (Schreiber, 1978), moths (Ricklefs and O'Rourke, 1975), ants (Kusnezov, 1957), and arthropod communities (Teraguchi et al., 1981) are most diverse in the tropics with diversity decreasing toward the poles. Within Alaska, the diversity of soil-dwelling mites decreases with increasing latitude (Thomas and MacLean, 1988). Braconid wasps (Quicke and Kruft, 1995), ichneumonid wasps (Sime and Brower, 1998), and bees (Michener, 1979), however, have peaks of diversity some distance from the tropics. Sawflies (Kouki et al., 1994), aphids (Dixon et al., 1987), and springtails (Rapoport, 1975) even display a reversal of the general trend, with highest diversities at relatively high latitudes.

Though many hypotheses have been proposed to explain the latitudinal diversity gradient (see Pianka, 1966; MacArthur, 1972; Rosenzweig, 1995; Mittelbach et al., 2007, for reviews), there is little agreement as to its causes (Rosenzweig, 1995; Lawton, 1999). Many of the hypotheses explaining the latitudinal diversity gradient depend on climatic trends. For example, Stevens (1989) believed that Rapoport's rule (the observation that the latitudinal extents of species' ranges are positively related to latitude) could be explained by the climatic tolerances of species. In the tropics, where microclimate of a site may vary little, a species with narrow climatic tolerances may persist. These narrow tolerances restrict it to occupying only a small range in which those particular climatic conditions are realized. A high latitude species, however, must be able to endure a large range of temperatures over the course of a year, so it must be able to survive under a broad range of temperatures. This wide tolerance predisposes it to being able to survive over a large geographic range. Stevens discussed multiple reasons why the latitudinal diversity gradient might be related to Rapoport's rule. In some of the other hypotheses explaining the latitudinal diversity gradient, differences in rates of speciation are often based on temperatures and lengths of growing seasons. The evolutionary and time hypotheses propounded by Mittelbach et al. (2007) rely on these kinds of underlying climatic gradients. For example, the length of time over which speciation has occurred is often defined by time since severe climatic events, especially glaciations. It is clear that climate characteristics are some of the most important determinants of the latitudinal diversity gradient (Whittaker et al., 2001; Grytnes and McCain, 2007).

Elevational Gradients Biodiversity generally decreases with elevation or peaks at middle elevations (Rahbek, 1995, 2005; Grytnes and McCain, 2007). Arthropod diversity also tends to be highest at low (Wolda, 1987; Fernandes and Price, 1988; McCoy, 1990; Kearns, 1992; Stevens, 1992; Olson, 1994; Sparrow et al., 1994) or intermediate elevations (Janzen, 1973; McCoy, 1990; Olson, 1994; Sanchez-Rodriguez and Baz, 1995; Fleishman et al., 1988; Sanders, 2002; Mac Nally et al., 2003). A few arthropod groups, such as tropical psocids and bees (Turner and Broadhead, 1974; Gauld, 1987), are most diverse at high elevations.

As with the latitudinal diversity gradient, multiple hypotheses have been proposed to explain observed elevation patterns (see McCoy, 1990; Lomolino, 2001; Rahbek, 1995; Grytnes and McCain, 2007, for reviews). Stevens (1992) and Lomolino (2001) noted that many of the

same hypothesized climatic determinants that result in the latitudinal diversity gradient are reproduced in miniature in mountain systems, so some of the same drivers may be causing both patterns. Price et al. (1998), in a study of global diversity patterns of gall-making insects, even considered the climatic influences of latitude and climate to be interchangeable by the equation, 4° latitude = 305m elevation. Lomolino (2001) predicted that, if climate causes the observed elevational patterns of diversity, then the elevational peaks of diversity should shift down slope with increasing distance from the equator. McCain (2007) proposed a model for species diversity of bats that predicted variation in the elevation of maximum diversity based on general gradients of temperature and water availability on mountains (Figure 1.1). Temperature generally decreases with elevation (Barry, 1992). Precipitation is generally highest at middle to upper elevations (Grytnes and McCain, 2007), but McCain (2007) noted that much of the precipitation received by the highest elevations falls as snow, which is less biologically available. Her model predicted a mid-elevation peak of diversity where lower elevations are arid and a monotonically decreasing diversity pattern from low to high elevations where lower elevations receive plentiful rainfall. If her model were to be applied over the latitudinal climatic gradient while precipitation was held constant, peak diversity would be expected to move down slope with increasing latitude as was predicted by Lomolino (2001) since temperatures generally decrease with latitude. Consistent with these predictions, McCoy (1990), in a meta-analysis of patterns of insect diversity along elevational gradients, found that the elevation of maximum diversity decreased with latitude, with the elevation of highest diversity dropping to near sea level at 60° latitude.

McCain (2007) did not explain why bat species diversity was positively related to temperatures and water availability except to note that bats require a minimum temperature for activity. Grytnes and McCain (2007) gave two biological explanations for this trend. The first was that the requirements of many taxa such as intolerance to climatic severity or a need for high moisture preclude the existence of many species in the extreme cold of the highest elevations and in arid lower elevations. Although they did not cite Wright (1983), the second reason given by Grytnes and McCain (2007) was an application of Wright's species-energy theory (described below). Since productivity is generally highest at low to middle elevations where the climate is both warm and wet, species richness should be highest in these same areas based on Wright's hypothesis.

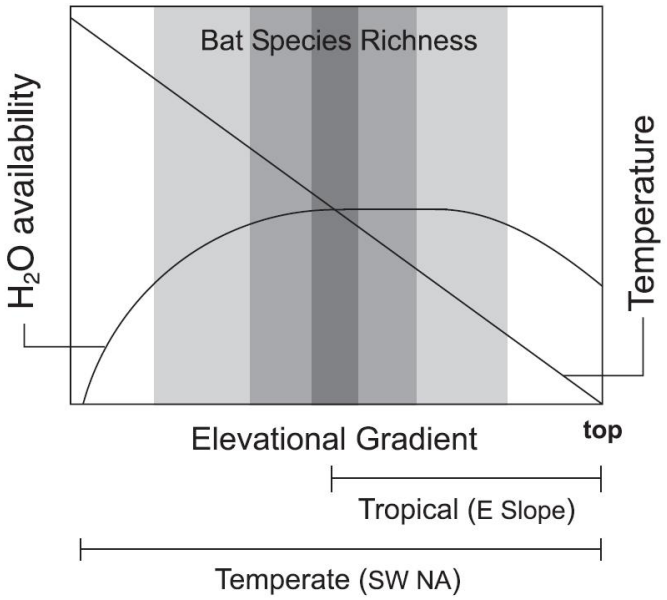

Figure 1.1: Generalized climatic model for elevational gradients in species richness of bats from McCain (2007). Bat species richness is depicted in grey tones with darker tones indicating more species. The placements of generalized tropical and temperate elevational gradients analyzed in McCain's study are shown below the x-axis.

Relationships of Diversity with Productivity

The most common relationship between biodiversity and productivity is unimodal, with maximum diversity at an intermediate level of productivity, but a monotonically increasing relationship is also common (Rosenzweig, 1995; Mittelbach et al., 2001). Diversity of terrestrial arthropods often increases monotonically with productivity (Brown and Davidson, 1977; Morton and Davidson, 1988; Seimann, 1998; Srivastava and Lawton, 1998; Kaspari et al., 2000; Bailey et al., 2004). Mittelbach et al. (2001), in a meta-analysis of studies of diversity, including 11 focused on terrestrial arthropods, found that unimodal relationships were the most common.

The species-energy theory of Wright (1983) explains well the relationship of increasing diversity with productivity in poor environments. It posits that when productivity is low, population sizes cannot be large, so extinction rates are high, resulting in the observed pattern. An explanation for why diversity often decreases after productivity exceeds a certain point is more contentious. Although Rosenzweig (1995) listed reservations about all of the nine hypotheses he reviewed explaining this unimodal pattern, he seemed to favor most the

hypothesis of Tilman (1982) that decreasing environmental heterogeneity at the high end of the productivity spectrum reduces the number of niche types, thereby reducing the number of species that can make use of them. The assumption of this hypothesis is that heterogeneity is highest somewhere in the middle of the productivity spectrum. This may be true when only a few species benefit from high productivity. Many species may not be able to take advantage of increased productivity, so they would not be able to respond to heterogeneity in productivity at the high end of the productivity spectrum. For example, Grime (1973*a*,*b*) found that at high productivities, a few highly competitive species of herbaceous plants reduced richness by excluding other species. As another example, fires may cause a unimodal pattern in vascular plant diversity in forests of the western Kenai Peninsula. Some of the most productive areas in this region are areas where young hardwoods (*Betula papyrifera* var. *kenaica* and *Populus tremuloides*) have established after fires. These hardwoods grow quickly and thickly in burns, out-competing most other plants, producing a habitat that is nearly a monoculture with little heterogeneity. Vascular plant diversity increases and productivity decreases as spruce trees (*Picea* spp.) succeed hardwoods and the forest becomes more heterogeneous. The species-area relationship of increasing diversity with area, which has more support than any other diversity pattern (Rosenzweig, 1995), is another proposed mechanism for the unimodal productivity relationship. Since the regions of highest productivity are usually smaller than less productive regions, diversity may be low in these high productivity regions simply due to the species-area relationship. It is also important to note that if productivity is measured as some biological attribute, such as dry weight of plants per hectare, evapotranspiration rate, or the density of chlorophyll, then the measured productivity is itself a product of characteristics of the environment.

Congruence of Diversity Patterns of Arthropods and Plants

The diversity of plants is a good predictor of insect species richness in natural systems (Gaston, 1992). In natural systems, herbivore diversity is generally correlated with plant diversity (Seimann, 1998). Many insects, especially herbivores, are tightly associated with particular plant species. For these specialized insects, existence depends on the presence of their host plants. An alternative explanation for a positive correlation between plant and insect diversity is that more plant species provide greater structural diversity and niches for

a greater number of insect species. Discerning between these two hypotheses can be difficult (Murdoch et al., 1972). In addition, the same environmental attributes that render some places hospitable and others harsh may act similarly on plants and insects, even among generalists that do not require particular plant species or structural characteristics. The Aphididae are an exception in which species diversity is inversely related to plant species diversity (see Dixon, 1985, for an explanation).

Practical Considerations in Measurement of Arthropod Diversity

Taxonomic Resolution

Danks (1996) strongly advocated insect identifications at the species level for two reasons: species are the functional entities in ecology, and biological information is associated with species names. When studying biodiversity, species level resolution or the even finer molecular resolution reveals the most information about a system (Lenat and Resh, 2001). Identifications at coarser taxonomic ranks such as genus, family, or order yield progressively less information. Nonetheless, since related species tend to have similar characteristics and requirements, broader taxa tend to show similar patterns of distribution and abundance as the species that compose them so that species-level information and coarser taxonomic information may be largely redundant (Bowman and Bailey, 1997; Pik et al., 1999; Waite et al., 2004; Marshall et al., 2006). Improving taxonomic resolution may not improve the resolution of the results (Oliver and Beattie, 1996).

Species identifications for terrestrial arthropod groups require a high level of expertise and are labor intensive, time consuming, and expensive (Danks, 1996). In some cases, specific determinations are not possible in the short term either because expertise is not available or because the taxonomy is yet to be resolved. Sorting to coarser taxonomic ranks does not require as much taxonomic knowledge and is faster and cheaper than identifications to species (Balmford et al., 1996). In addition, fewer misidentifications tend to be made at coarser taxonomic resolutions, particularly by workers with less than the highest level of expertise (Marshall et al., 2006). Because of the cost differences and the redundancy of information gained at nested taxonomic levels, the optimal taxonomic resolution for a study may be broader than the species level (Oliver and Beattie, 1996; Paoletti and Bressan, 1996;

Bailey et al., 2001). The appropriate resolution depends on the objectives and budget of an investigation.

In a lotic system, Marshall et al. (2006) found that the most efficient taxonomic resolution for detecting differences among communities was the family level. In their study, the cost of species identifications was over six times higher than family level identifications, but using family data resulted in only a 6% loss in information compared with species identifications in terms of detecting differences between sites (Figure 1.2). In a terrestrial system, genus resolution was found to be optimal for distinquishing between sites and predicting species richness of beetles and moths (Mandelik et al., 2007). A consensus among several studies on arthropods and plants (Balmford et al., 1996; Bowman and Bailey, 1997; Marshall et al., 2006; Mandelik et al., 2007) is that the family or genus resolutions are optimal in terms of efficiency. Due to the especially high effort, cost, and time required for identification of terrestrial arthropods, it may be appropriate to focus identifications at the family level.

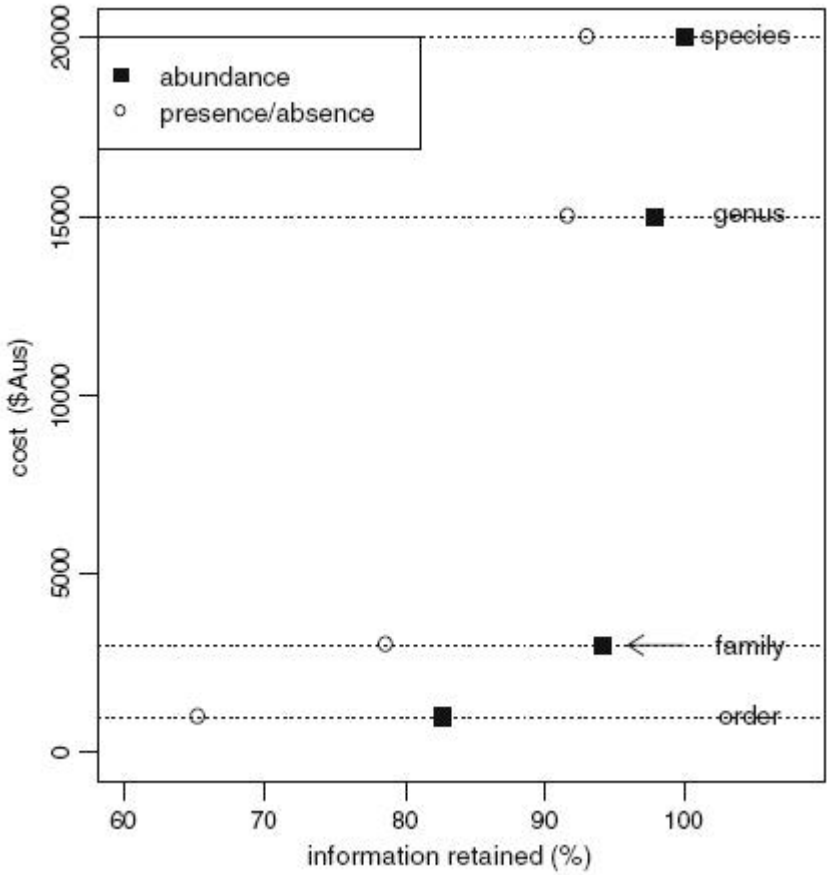

Figure 1.2: Information retained vs. cost of obtaining taxonomic data. The arrow represents the optimum cost/benefit relationship. Figure from Marshall et al. (2006).

Taxonomic Breadth

For biodiversity studies, Kremen et al. (1993) recommended sampling several higher taxa of arthropods using standardized methods. The criteria they gave for choosing groups were that several distantly related groups should be sampled and each group should be diverse with high endemism in the sampled area. An argument for including a wide taxonomic scope is that different groups of arthropods in the same area may exhibit quite different patterns of diversity (e.g., Prendergast et al., 1993; Wettstein and Schmid, 1999). Danks (1996) acknowledged that the selection of taxa for inclusion in a biodiversity study is usually a compromise between feasibility and the objectives of the study. He also cautioned against studying only the most easily identified groups, since these may provide little useful information toward the objective.

Within a fixed budget, a choice between wide taxonomic breadth and fine taxonomic resolution may be necessitated. Since different groups often exhibit distinct patterns of diversity and because conclusions at differing taxonomic levels are generally concordant, sacrificing taxonomic resolution is probably more optimal than narrowing taxonomic scope in diversity studies that are intended to be generalizable (Heino et al., 2003).

Objectives and Hypotheses

To contribute to the understanding of arthropod diversity and provide information relevant to monitoring and conservation, I examined patterns of terrestrial arthropod diversity on the Kenai National Wildlife Refuge (KENWR). The primary goals of this project were to document patterns of terrestrial arthropod diversity in a boreal forest/tundra system and explain the causes of the observed patterns. Specifically, I attempted to elucidate the relationships of arthropod family diversity with elevation, local climate, productivity, and vascular plant species density.

Objective 1: Document patterns of terrestrial arthropod diversity.

I summarized observed patterns of arthropod family density and produced a map of predicted arthropod family density over the KENWR. These data alone will be useful to the KENWR for inventory and management purposes.

Objective 2: Explain the causes of the observed pattern of diversity.

Hypothesis 1: I hypothesized that, at the high latitude of the KENWR, cold temperatures limit the number of arthropod families. I predicted that, as a result, indices of arthropod diversity are directly related to temperature.

Hypothesis 2: I hypothesized that, over the relatively dry landscape of the KENWR, low water availability excludes some arthropod taxa from drier sites. I predicted that diversity indices are positively related to precipitation.

Hypothesis 3: Because of increasing climatic severity (temperature, wind, and extreme variation) with elevation, I hypothesized that fewer arthropods can persist at higher elevations

than at the milder climates of lower elevations. I predicted that arthropod diversity indices decrease with elevation monotonically.

Hypothesis 4: I hypothesized that generally low productivity on the landscape of the KENWR limits arthropod density and diversity. While productivity is negatively correlated with elevation on the KENWR, it varies widely within elevational bands due to local differences in soils, hydrology, and disturbance history; therefore it may explain differences in arthropod diversity not attributable to elevational gradients. I predicted that arthropod diversity indices monotonically increase with a remotely-sensed vegetation index of productivity.

Hypothesis 5: Particularly among herbivorous arthropods, geographic distributions are limited by the availability of their host plants. As a result, I predicted that arthropod diversity indices are positively related to the species density of vascular plants.

1.2 Methods

Overview

Most data used in this analysis are from the KENWR's Long-Term Ecological Monitoring Program (LTEMP), a grid-based multi-taxa inventory and monitoring effort on the KENWR. Specimens from all points were sorted to families, arthropod family diversity indices for each grid point were calculated, and hypothesized relationships of arthropod diversity to elevation, productivity, habitat structure, and plant diversity were tested using regressions. The spatial scale of this study was a 100 m^2 plot. The area of inference and geographic extent was the 800,000 ha KENWR excluding water bodies, glaciers, beaches, coastal bluffs, saline mud flats, and areas of human development. This study was limited to an analysis of alpha diversity at the family level of resolution.

Study Area

General Description

The 800,000 ha KENWR covers much of the western Kenai Peninsula in south-central Alaska (Figure 1.3). The KENWR is a vast, diverse landscape of boreal forest, alpine tundra,

and coastal marsh biomes. The western portion of the refuge is mostly lowland white spruce (*Picea glauca*), black spruce (*Picea mariana*), paper birch (*Betula papyrifera* var. *kenaica*), and quaking aspen (*Populus tremuloides*) forest; black spruce muskeg; graminoid and bryophyte-dominated wetlands; and lakes (Figure 1.4). Frequent fires on the lowlands have left a patchy mosaic of varying stand ages. The Kenai Mountains extend on a north-south line along the eastern portion of the refuge. This region comprises intervening forested valleys of spruce and western hemlock (*Tsuga mertensiana*); alpine tundra; barren, rocky mountain peaks; snowfields; and glaciers. Chickaloon Flats on the northern coast is the only tidal marsh on the refuge.

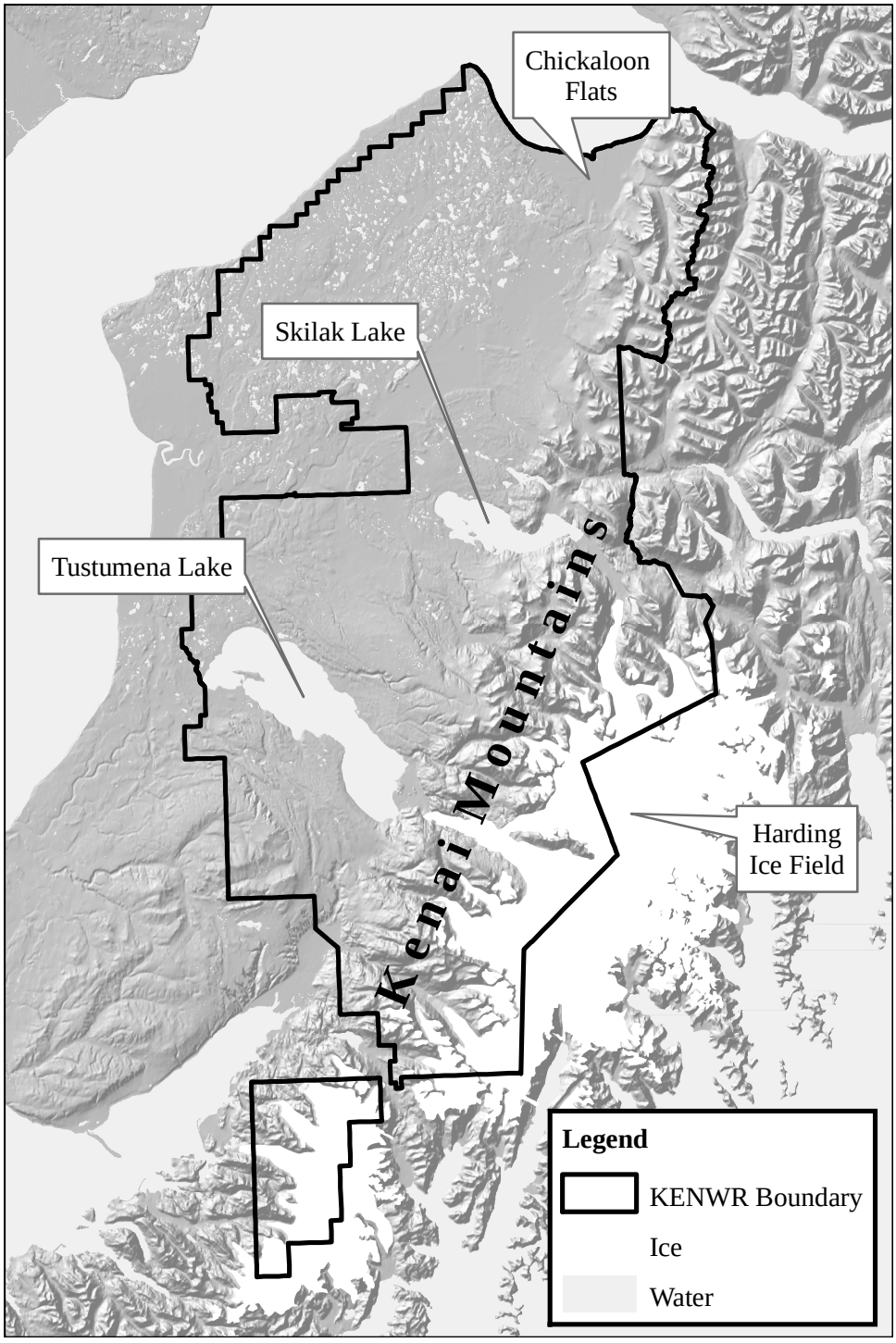

Figure 1.3: Map of the Kenai National Wildlife Refuge.

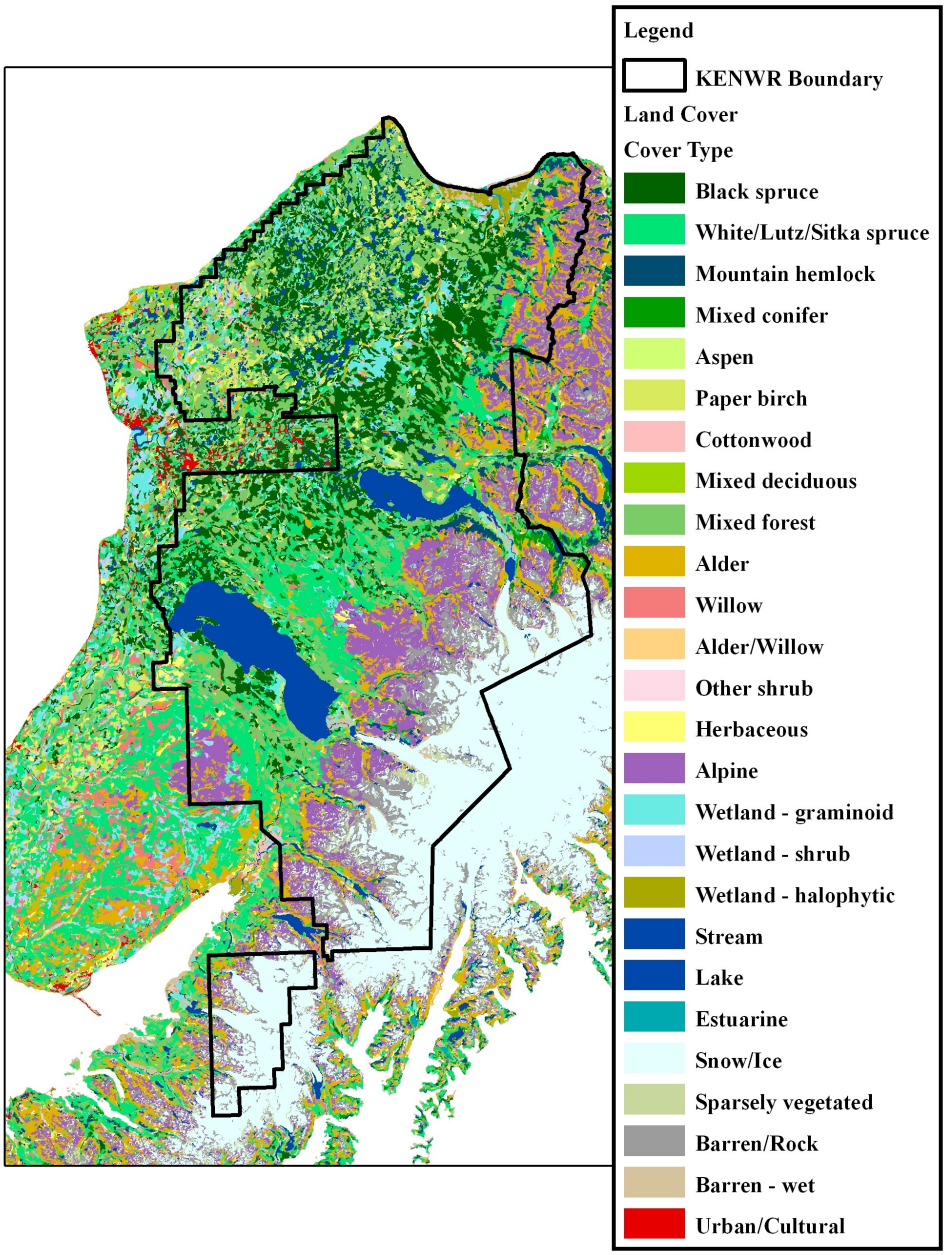

Figure 1.4: Kenai Peninsula land cover classification (O'Brien, 2006).

Climate

The KENWR, ranging from 59.4–61° north latitude and 149.7–151.2° west longitude, is large enough to exhibit distinct gradients of climate over the landscape. The vast lowlands of the northwestern portion of the refuge approach the high-latitude continental climate of Interior Alaska. Mountain ranges to the west, north, and south leave this region in a

rain shadow of low precipitation. Its high continentality also means that winter minimum temperatures are colder and maximum summer temperatures are hotter in this region than in other low elevation portions of the refuge. From here southward along the lowlands, precipitation increases and temperatures ameliorate until a maritime climate is reached at the extreme south of KENWR. In the Kenai Mountains running down the eastern side of the refuge, temperatures decrease and precipitation increases with elevation. North-south and west-east gradients of decreasing continentality also exist within the mountains, although the west-to-east gradient within the mountains is not clearly discernible within the narrow band of mountains on the refuge, becoming more clearly evident when the rest of the Kenai Mountains in the Chugach National Forest and Kenai Fjords National Park are included.

Introduction to LTEMP

LTEMP was designed to be a multi-taxa inventory and monitoring program for breeding birds, vascular plants, non-vascular plants, and terrestrial arthropods over the KENWR. A grid sampling design covering the entire refuge provided good spatial coverage and ensured that the sample was representative of the refuge as a whole. A quick, swat team-like approach to sampling was employed, sampling all taxa in less than one hour on each plot so that five to six plots could be visited per team per day. Collaboration with the USDA Forest Inventory and Analysis (FIA) program, which was already sampling vegetation characteristics across the KENWR, resulted in greater efficiency because redundant vegetation sampling was not necessary. The three goals of LTEMP were to (1) inventory current diversity using generalized sampling methods, (2) set the stage for statistical modelling of various attributes by sampling multiple metrics at the same points, and (3) initiate long-term monitoring of multiple taxa.

Sampling Design

FIA sampling in Alaska was performed by the USDA's Pacific Northwest Research Station (PNRS). A regular grid of points at 4.8km intervals was imposed over the south central Alaska region, 327 of which fell on the KENWR. Based on aerial photography and satellite imagery, all points that fell on land were classified as forested or non-forested.

Vegetation Sampling

FIA Vegetation Sampling Between 1999 and 2002, The Forest Service sampled vegetation using FIA protocols on the 176 points on the KENWR which had been determined to be forested. At each sampling site, a circular, 5.4m radius ($100m^2$) horizontal-vertical (HV) plot was established, where horizontal and vertical distributions of vegetation layers and their species composition were described. The observer divided the vegetation strata by natural breaks and recorded the density of each species present in each layer by ocular estimation of percent cover. All vegetation except for trees of over 2.5cm DBH (diameter at breast height) was included. See USDA Forest Service (2000-2003) for complete FIA protocols.

LTEMP Vegetation Sampling In 2004 and 2006, vegetation was sampled on the remaining 80 non-forested points on the KENWR by refuge staff using the LTEMP vegetation protocols. In order to complete a consistent vegetation presence/absence dataset across all LTEMP points, all species within a 5.4m radius plot centered on the LTEMP point were recorded.

After collapsing the FIA's 3-dimensional HV data into a simple list of species present within the 5.64m radius circular plot, the vegetative inventory information from both vegetation sampling methods was combined.

Plot Weather Data

Temperature (°C), relative humidity(%), wind speed (m/s), and time of day were measured on the plots immediately before athropod sweep net samples were taken. Data were collected using a Kestrel 3000 Pocket Weather™ meter and averaged over 30 seconds. The weather was also scored categorically using the codes and sky conditions descriptions in Table 1.1. The code numbers correspond to generally decreasingly favorable conditions for collection of sweep net samples from clear conditions to snow and rain showers.

Table 1.1: Sky condition codes and definitions.

Code	Conditions
0	Clear or a few clouds
1	Partly cloudy (scattered) or variable sky
2	Cloudy (broken) or overcast
4	Fog or smoke
5	Drizzle
7	Snow
8	Showers

Arthropod Sampling

Terrestrial arthropods were sampled in 2004 and 2006 at all points where vegetation data were collected ($n = 255$). All points were accessed by helicopter. KENWR staff sampled half of the grid in 2004 by visiting every other point ($n_{2004} = 152$); the second half of the points ($n_{2006} = 103$) were sampled in 2006 (Figure 1.5). Concurrent sampling of birds determined the seasonal (June 7–30) and daily (04:40 to 10:54 hours) sampling windows. In both years, terrestrial arthropods were sampled over the last three weeks in June. Points that fell on water, mud flats, or ice were not visited. Some points that were too far away from landing sites ($>$ 1km) or inaccessible due to steepness of terrain were also dropped.

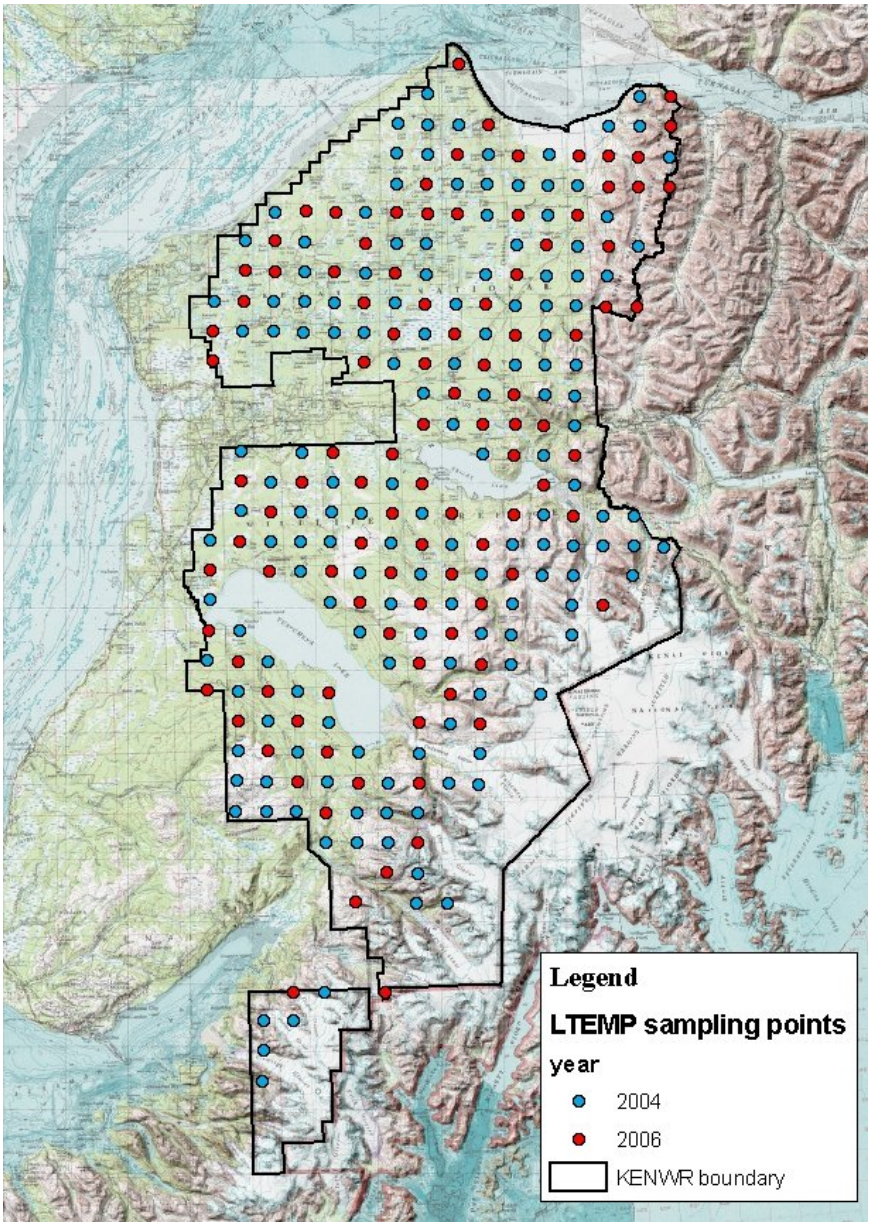

Figure 1.5: LTEMP sampling points. Blue points were sampled in 2004 and red points were sampled in 2006.

Arthropod Sampling Methods

Sweep netting was chosen as the sampling method for terrestrial arthropods because it performed well compared to other methods for broadly sampling the insect community, could be executed quickly, and caused little damage to the plots. A single sweep net sample was taken at each plot. For each sample, the collector swept a 30.5cm diameter aerial insect

net quickly back and forth over all vegetation and other substrates within reach over the entire 100m^2 circular plot within five minutes. The sample was then emptied into a vial of 80–90% ethanol. Care was taken to minimize damage to the arthropods and to prevent their escape during transfer from the net to the vial.

Taxonomic Resolution and Breadth

This study was restricted to examination of arthropod diversity at the family resolution. Taxonomic breadth was as broad as possible, although some groups were excluded due to practical constraints. Mites, spiders, Trichoptera, Microlepidoptera, and most Nematocera were excluded due to the inordinate time required to identify them. All Psocoptera were immature at the time of sampling, so they could not be identified. Lepidoptera and minute Nematocera were excluded because they tended to be badly damaged by the selected collecting methods. Terrestrial Gastropoda, the only other animals collected, were ignored. My analyses included 90 families (Table 1.2) and, except where specified, all further discussion will be limited to this set of families.

Sorting and Identification

All arthropod specimens were processed and archived using appropriate curation methods. All arthropods present in the samples were sorted into orders and most were sorted to families. Borror et al. (1989); Borror and White (1970); White (1983); McAlpine et al. (1981, 1987); Goulet and Huber (1993) were the main keys used for family level identifications. With the exception of small amounts of material lent to various systematists, all specimens remain in the arthropod collection of the KENWR (international collection coden: KNWR).

GIS-derived Data

Elevation Data

Elevation data were extracted from a digital elevation model (DEM) of the Kenai Peninsula (O'Brien, 2006). This was done by resampling USGS DEM data from 30m $\times$ 60m to 30m $\times$ 30m resolution.

Table 1.2: Arthropod families included in analyses.

Arachnida	Hemerobiidae	Otitidae
Opiliones	**Coleoptera**	Phoridae
Sclerosomatidae	Anobiidae	Pipunculidae
Chilopoda	Cantharidae	Psilidae
Lithobiomorpha	Carabidae	Rhagionidae
Lithobiidae	Chrysomelidae	Scathophagidae
Parainsecta	Coccinellidae	Sciomyzidae
Collembola	Curculionidae	Sepsidae
Entomobryidae	Elateridae	Simuliidae
Hypogastruridae	Lathridiidae	Sphaeroceridae
Isotomidae	Leiodidae	Stratiomyidae
Sminthuridae	Lycidae	Syrphidae
Insecta	Pythidae	Tabanidae
Odonata	Scarabaeidae	Tachinidae
Coenagrionidae	Scirtidae	Tephritidae
Orthoptera	Staphylinidae	**Hymenoptera**
Acrididae	**Diptera**	Aphelinidae
Hemiptera	Agromyzidae	Apidae
Acanthosomatidae	Anisopodidae	Argidae
Achilidae	Anthomyiidae	Bethylidae
Anthocoridae	Anthomyzidae	Braconidae
Aphididae	Asteiidae	Ceraphronidae
Cicadellidae	Bibionidae	Diapriidae
Delphacidae	Chamaemyiidae	Dryinidae
Lygaeidae	Chloropidae	Encyrtidae
Miridae	Clusiidae	Eulophidae
Nabidae	Culicidae	Eurytomidae
Psyllidae	Dolichopodidae	Formicidae
Thysanoptera	Drosophilidae	Ichneumonidae
Phlaeothripidae	Dryomyzidae	Platygasteridae
Thripidae	Empididae	Pteromalidae
Plecoptera	Ephydridae	Scelionidae
Chloroperlidae	Heleomyzidae	Tenthredinidae
Neuroptera	Lauxaniidae	Torymidae
Chrysopidae	Micropezidae	
Coniopterygidae	Muscidae	

Productivity Data

The Normalized Difference Vegetation Index (NDVI) is a widely used, remotely sensed estimator of primary productivity in terrestrial systems. The NDVI makes use of the difference between the high absorbance of chlorophyll in the red portion of the electromagnetic spectrum and its low absorbance in the near-infrared band to measure the density of visible chlorophyll in an area. NDVI data are positively correlated with primary productivity (Box et al., 1989; Lo Seen Chong et al., 1993; Ricotta and Avena, 1998), the density of active photosynthetic pigments in vegetation (Myneni et al., 1995), actual evapotranspiration (AET) (Box et al., 1989; Lo Seen Chong et al., 1993), and leaf area index (LAI) (Soudani et al., 2006). I calculated NDVI (Figure 1.6) from a mosaic of LandSat 7 imagery covering the Kenai Peninsula (O'Brien, 2006). The original imagery was a set of LandSat 7 images taken in 2002 available from the Multi-Resolution Land Characteristics Consortium, which covered the area of interest at 30m resolution.

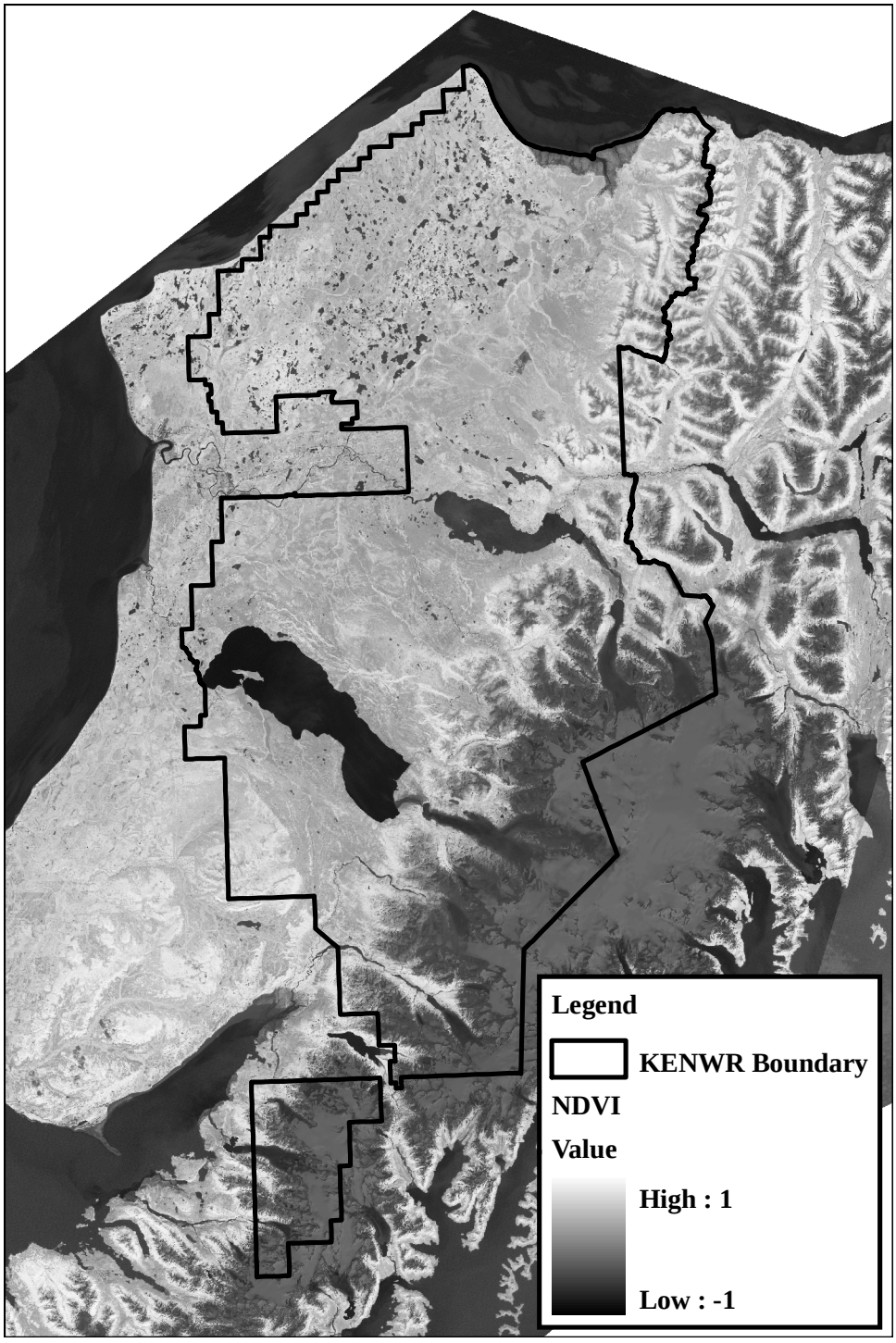

Figure 1.6: NDVI raster generated from LandSat 7 imagery.

Climate data

Precipitation and temperature data were extracted from the PRISM (Parameter-elevation Regressions on Independent Slopes Model) raster datasets produced by the Spatial Climate Analysis Service at Oregon State University (SCAS/OSU). This dataset is currently the best climate coverage available for Alaska (Simpson et al., 2005). The model incorporated climatological point data for the time period of 1961-1990 and a digital elevation model (DEM) to interpolate monthly precipitation and temperatures over the state of Alaska at 2km resolution.

Rarefaction

In order to assess how the sweep net method sampled the arthropod fauna, a rarefaction curve was generated using the `specaccum` function of the `vegan` package (Oksanen et al., 2007).

Prediction by Random Forest Regression

In order to create a continuous raster of arthropod family density over the KENWR, predictions were made by Random Forest Regression, a machine learning algorithm with high predictive accuracy (Breiman, 2001).

A 100m $\times$ 100m resolution prediction grid was imposed over the KENWR. Values of covariates that were available as raster datasets were extracted by resampling (bilinear interpolation) using ArcMap (Table 1.3). Covariates available as vector datasets were converted to raster datasets conforming to the prediction grid. For all covariates, values from pixels in which LTEMP sampling locations fell were extracted for use in fitting of random forest regression models. For prediction, the time was set at median observed values from observed LTEMP data (June 18 at 8:00 am).

Random forest regressions were fitted using the `radomForest` function of the `randomForest` package (Liaw and Wiener, 2002). For all random forest regressions, 5,000 trees were built. To select an optimal value of the 'mtry' parameter, the number of predictors randomly selected for consideration at each node, I ran random forest regressions for all values of mtry between 1 and 50 and selected the value of mtry that yielded the highest value of pseudo-R-squared and the lowest mean squared error. A random forest regression model was then

Table 1.3: Covariates included in random forest regressions.

Coefficient	No.	Description
Spatial		
latitude	1	
longitude	1	
Topographic		
elevation	1	Elevation from the Digital Elevation Model (DEM). See page 20.
slope	1	Derived from DEM.
aspect	1	Derived from DEM.
curvature	1	Derived from DEM.
distance_ocean	1	Distance to the ocean.
Climate		
precipitation	13	monthly and annual precipitation from the PRISM model. See page 24.
temperature	13	average monthly and annual temperature from the PRISM model. See page 24.
accumulation	1	Amount of accumulated surface runoff based on PRISM annual precipitation data and DEM.
Vegetation		
NDVI	1	See page 22.
landcover	1	Vegetation cover classes from O'Brien (2006).
Temporal		
day	1	Julian day.
hour	1	hours since midnight.
Historic		
years_post_fire	1	Years since last fire.
Total	39	

fitted using this optimal value of the mtry parameter. I used this model to make predictions at all 100m $\times$ 100m pixels over the KENWR.

Diversity Indices

Selection and Calculation of Diversity Indices Because richness and dominance/evenness are quite different qualities of diversity, three diversity indices were calculated to represent the spectrum from richness-only to dominance/evenness-only measures of diversity. Family density (S_i) was simply the number of families collected at plot i, an index of only richness information. Shannon's Information Index (H_i'), a compsite index reflecting richness and

evenness (but more richness than evenness) was calculated as in Magurran (1988) using equation 1.1

$$H_i' = -\sum_{j=1} q_{ij} \ln(q_{ij}) \tag{1.1}$$

where q_{ij} is the proportion of individuals from the j^{th} family (estimated by m_{ij}/N_i, where m_{ij} is the number of individuals of family j collected at plot i and N_i is the total number of specimens collected at plot i). The inverse of the Berger-Parker index, hereafter referred to as the Berger-Parker index (d), measures evenness and was calculated as in Magurran (1988) using equation 1.2

$$d_i = N_i / m_{i\max} \tag{1.2}$$

where N_i is the total number of families collected at plot i and $m_{i\max}$ is the number of individuals collected from the most abundant family at plot i. The inverse of the Berger-Parker index was used so that it would increase with increaseing evenness, consistent with other diversity indices. It is a dominance-only index containing no information about richness. From among the list of diversity indices that have been devised, family richness and the Berger-Parker index were chosen because they are easily interpreted measures of distinct aspects of diversity. Of the composite indices, which are more difficult to interpret, Shannon's Information Index was selected because it performs better than most other composite indices in terms of distinguishing differences between plots (Taylor, 1978; Kempton, 1979) and because of mathematical issues encountered for some other diversity indices at the often small sample sizes observed within individual plots.

Influences of Sample Size and Detection Probabilities When comparing diversity indices, it is important to be aware of some of their properties. Both S and H' are influenced by the number of specimens collected (Magurran, 1988). The number of specimens collected at a site ($N_{i\text{ observed}}$) is directly related to detection probabilites by the equation

$$N_{i\text{ observed}} = \sum_{i=1}^{n} N_{ij} p_{ij} \tag{1.3}$$

where N_{ij} is the true number of individuals of family j present at site i and p_{ij} is probability of detecting family j at site i. Note that there is no way to obtain good estimates of N_{ij} without more information (i.e., some way of estimating the p_{ij}'s). In a similar way, observed values of S are directly related to detection probabilities by the equation

$$\bar{S}_i = \sum_{i=1}^{n} P_{ij} p_{ij} \tag{1.4}$$

where $\bar{S}_i$ is the expected value of observed richness at site i, P_{ij} is the probability that family j is present at site i, and p_{ij} is the probability of detection of family j at site i given that it is present. Note that the term p_{ij} had two different definitions in equations 1.3 and 1.4. The point is that the regressions of S and H' reflect not just richness, but also abundance and detection probabilities.

Calculability of Diversity Indices Another point which must be taken into account is that H' and d could not be calculated at all points. d cannot be calculated when $S = 0$ and H' cannot be calculated when $S = 0$ or 1. As a result, H' and d do not apply to the sites with lowest diversity. In contrast, S can be calculated everywhere, so regressions of S contain information both about family richness and the distribution of arthropods (at least whether or not athropods are detected).

Analyses of Spatial Autocorrelation

I assessed spatial autocorrelation of diversity indices by visual inspection of empirical variograms generated with the `variog` function in the `geoR` package (Ribeiro and Diggle, 2001). Inverse-distance-weighted Moran's I tests using the `moran.test` function in the `spdep` package (Bivand et al., 2007) were performed to test for spatial autocorrelation. In assigning weights for the Moran's I tests, the maximum distance for comparison of points was set at 50km because greater distances were much more computationally intensive and did not substantially alter results.

Regression Analyses of Diversity Indices

Two of the 255 plots sampled were excluded from regression analyses, leaving 253 plots for the analyses. In one case, plot vegetation data had been misaligned with all other data due

to logistical errors. In another case, it appeared that an avalanche had cleared the plot after NDVI and plot vegetation data were collected. The site had been lush and vegetatively diverse when NDVI and vegetation data were observed. Later, when the sweep net sample was taken, nearly all vegetation had been scoured away, leaving mostly bare earth. No other sites appeared to have experienced such drastic changes between the times when various data were collected.

All regression models were fitted with appropriate generalized linear models (GLMs). Numerous examples exist of regressions of count data, including models of richness, but it appears that regressions of other diversity indices are less common. Benin et al. (2003) used Gaussian ordinarly least squares (OLS) regressions for modelling values of Shannon's Information Index (H') and the Berger-Parker index (d). I selected optimal model families and link functions on the basis of the observed distributions of the diversity indices and by comparing measures of goodness-of-fit of plausible model families as in Zeileis et al. (2007). Family density (S) was modeled best by a negative binomial model fitted with the `glm.nb` function of the `MASS` package (Venables and Ripley, 2002). OLS regression was determined to be appropriate for Shannon's Information Index (H'), fitted with the `lm` function of the R `base` package. Since H' could not be calculated when S was 0 or 1, these values were excluded. As a result, only 226 records could be used in OLS regressions of H'. The (inverse of the) Berger-Parker index (d) was best modeled by gamma regression using an inverse link and fitted with the `glm` funciton of R `base`. (I did also model the untransformed Berger-Parker index with OLS regression as was done by Benin et al. (2003), which yielded essentially identical results and slightly poorer model fit.) Since the Berger-Parker index could not be calculated where no arthropods were collected, these records were omitted from analysis, leaving 240 records. Benin et al. (2003) omitted undefined values of H' and d in the same way.

Bayesian Model Averaging (BMA) was employed for model selection as well as for obtaining parameter estimates. Compared to other model selection methods based on AIC, BMA is more conservative, generally selecting simpler models (Wintle et al., 2003). BMA has the additional advantage of taking into account uncertainty due to model structure when estimating parameter values and their variances (Wintle et al., 2003). I performed BMA analyses using the `bic.glm` and `bicreg` functions of the `BMA` package (Raftery et al.,

2006). A large number of variables associated with the arthropod sweep net samples as well as selected interaction terms were included in BMA analyses (Table 1.4). I included only interaction terms that seemed most relevant to the hypotheses and predictions under consideration. The model most favored by the BMA analysis of each diversity index was selected as the best model. For the negative binomial BMA of S, where the dispersion parameter could not be fitted by the `bic.glm` function, a value for the dispersion parameter was chosen interactively by first running the BMA analysis, then fitting the best model. The dispersion parameter fitted by the best model was then used in another BMA analysis. In this case, the second BMA analysis selected the same best model. For each of the diversity indices, the best model selected by the BMA anlaysis was then fitted using GLM or OLS regression.

Table 1.4: Terms included in BMA Analyses.

Coefficient	Description
First-order terms	
E	Elevation from the Digital Elevation Model (DEM). See page 20.
Slope	Slope in degrees. Derived from DEM.
Aspect	Aspect. Derived from DEM. Aspects were converted from degrees (yielded by GIS conversion from the DEM) to values from -1 to 1 by the function $Aspect = -\cos\left(Aspect(^\circ) \times \frac{\pi}{180^\circ}\right)$ so that north-facing slopes were given a value of -1, east-facing slopes and west-facing slopes were given values of 0, south-facing slopes were given values of 1, and other aspects were assigned intermediate values. Level areas were assigned a value of 0.
Area	Area of elevational bands over the Kenai NWR. Calculated from DEM.
NDVI	See page 22.
P_{A}	Total annual precipitation from the PRISM model. See page 24.
T_{A}	Average annual temperatures from the PRISM model.
T_{O}	Observed plot temperatures at the time sweep net samples were taken. See page 17.
Humidity	Observed relative humidity. See page 17.
Wind	Observed wind speed. See page 17.
Sky	Observed sky condition. See page 17.

continued on next page...

Table 1.4 continued.

Coefficient	Description
Day	Julian day when sweep net samples were taken.
$Hour$	The time of day at which sweep net samples were taken.
$h_{(\text{cover classes})}$	Vegetation cover classes (O'Brien, 2006, Figure 1.4). The cover classes in which the sweep net samples fell were the following: hemlock, alpine, snow, white spruce, mixed conifer, paper birch, cottonwood, mixed deciduous, mixed forest, alder, willow, alder-willow, herbaceous, graminoid wetland, shrub wetland, halophytic wetland, stream, sparse and barren.
S_{veg}	Vegetation species density within the 5.4m radius plot. See page 17.
Climate interactions E:T_{A}, E:P_{A}, E:T_{O}, T_{A}:P_{A}, T_{A}:T_{O}, P_{A}:T_{O}	
Productivity interactions $NDVI$:T_{O}, $NDVI$:T_{A}, $NDVI$:P_{A}, $NDVI$:E, $NDVI$:S_{veg}	
Differences in seasonal phenology E:day	
Temporal changes in temperature T_{O}:Day, T_{O}:$Hour$	

Conventional R^2 measures of goodness-of-fit are not appropriate for generalized linear models such as negative binomial and gamma models (Cameron and Windmeijer, 1996), so for these models a deviance measure (R^2_{dev}) was calculated by the equation

$$R^{2*}_{\text{dev}} = 1 - \frac{\text{residual deviance}}{\text{null deviance}} \tag{1.5}$$

In a similar way, ANOVA and Analysis of Deviance sums of squares for each coefficient were expressed as a kind of standardized sum of squares in the same scale as R^2 and R^2_{dev}. These were expressed as R^{2*}_a and $R^{2*}_{\text{dev}a}$ for OLS and GLM models, respectively, by the equations

$$R^{2*}_a = \frac{\text{SS}_a}{\text{SS}_{\text{Total}}} \tag{1.6}$$

and

$$R^{2*}_{\text{dev}a} = \frac{\text{deviance}_a}{\text{null deviance}} \tag{1.7}$$

where SS_a was the sum of squares of coefficient a, SS_{Total} was the total sum of squares, and $deviance_a$ was the deviance explained by coefficient a.

Regressions of Observed Temperatures and NDVI

In order to better interpret relationships of diversity indices with observed plot temperatures (T_O) and NDVI ($NDVI$), OLS regressions were performed and a best model was selected using BMA. T_O was regressed against E, T_A, Day, and $Hour$; $NDVI$ was regressed against E, T_A, P_A, $Slope$, and $Aspect$. In both cases, all second-order interaction terms were also included.

1.3 Results

Quantity of Material Collected

Numbers of Specimens The 255 sweep net samples yielded a total of 15,136 specimens, of which 20 were terrestrial gastropods and the remaining 15,116 were arthropods. The total number of specimens collected per sample (N_T) ranged from 0 to 462 with a median value of 46 (Figure 1.7). The highest numbers of specimens per sample were collected in coastal halophytic wetlands and at lowland forested sites, particularly sites with open forest and a lush herbaceous understory. Progressively fewer specimens were collected with rising elevation until often no specimens were collected at high elevation, rocky sites and high tundra sites that had only recently become free of snow. No specimens were encountered on snowfields. The 90 families selected for analysis comprised 9,961 specimens. Of these, the number of specimens per sample (N) ranged from 0 to 309 with a median value of 27.

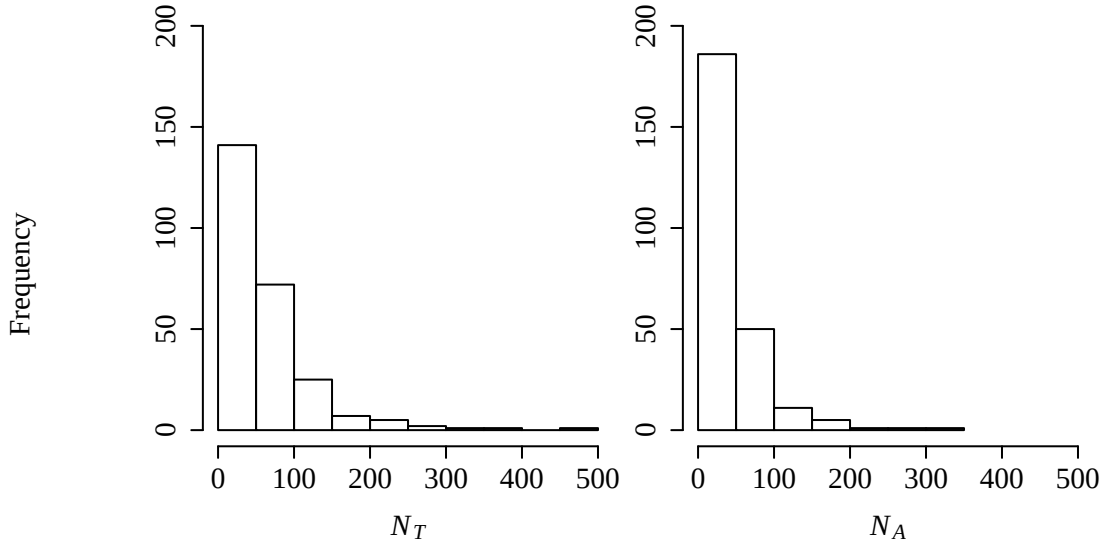

Figure 1.7: Numbers of specimens collected per plot. N_T is the total number of specimens collected per plot and N_A is the number of specimens per plot included in the analyses.

Taxa Collected

Orders Seventeen orders of arthropods were represented in the sweep net samples. Over half of the total specimens collected (53%) were Diptera. Hemiptera (20%), Araneae (9.0%), Collembola (7.0%), and Hymenoptera (4.5%) comprised substantial fractions of the total. Acarina (2.6%), Lepidoptera (1.1%), Coleoptera (0.94%), Thysanoptera (0.60%), Psocoptera(0.42%), Neuroptera (0.19%), Opiliones (0.15%), Orthoptera (0.086%), Trichoptera (0.026%), Odonata(0.020%), Plecoptera (0.020%), and Lithobiomorpha (0.013%) were less abundant in the samples.

Families The abundance and frequency of specimens from each family greatly varied (Table B.1). Culicidae, with a total (m_j) of 3,697 individuals collected and a frequency (f_j) of 0.76 (i.e., collected at 76% of sites), was the most abundant and frequently collected family. Aphididae, Sminthuridae, Cicadellidae, Muscidae, Delphacidae, Empididae, Ichneumonidae, Simuliidae, Braconidae, Lauxaniidae, Phoridae, Biobionidae, Anthomyiidae, and Cantharidae were also relatively abundant and frequently collected. Ephydridae were abundant locally, one site on the margin of Chickaloon Flats yielding 133 of the 135 Ephydrid specimens collected; but they were infrequent, collected at only three sites ($f_j = 0.012$). In contrast,

Diapriidae were relatively common (collected at 27 sites, $f_j = 0.11$) but were usually represented by few individuals at each site ($m_j = 38$, an average of 1.4 specimens per site where they occurred). Many taxa were rarely encountered and were represented by few specimens. For 17 families (Isotomidae, Chloroperlidae, Achilidae, Nabidae, Phlaeothripidae, Anobiidae, Carabidae, Lathridiidae, Lycidae, Pythidae, Scarabaeidae, Scirtidae, Clusiidae, Dryomyzidae, Sepsidae, Aphelinidae, and Bethylidae), only a single specimen was collected.

A rarefaction curve (Figure 1.8) of the family data shows a rapid rise in the number of families represented with increases in the number of sampling sites up to around 30. After about 40 samples, new families were added more slowly. Although most families were represented by the first 100 samples, an asymptote had not been reached by 255 samples.

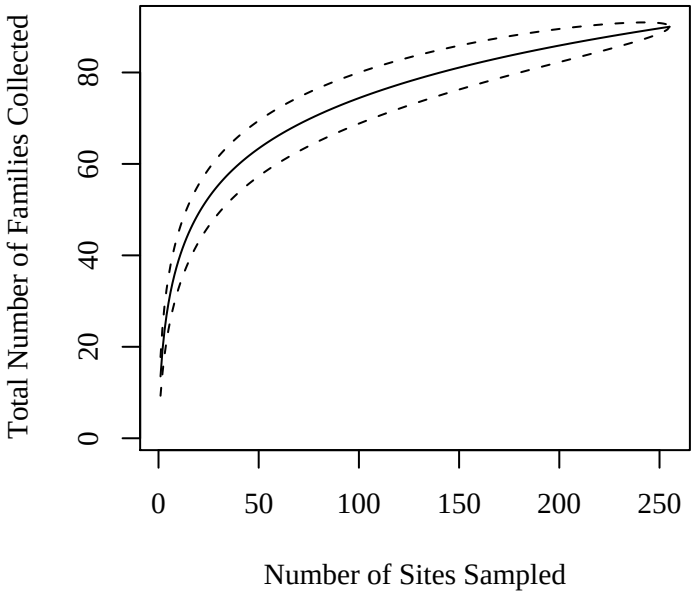

Figure 1.8: Rarefaction curve of families collected The solid line is the predicted number of families accumulated and the dashed line bound 95% confidence intervals.

Diversity

The three diversity indices ranged from family density (S) that accounts only for the number of taxa (richness) to the Berger-Parker index (d) that reflects only dominance. The

Shannon's Information Index (H') is intermediate, reflecting both richness and evenness, but it is influenced by richness more than evenness (Magurran, 1988).

Family density (S) was generally low, varying from 0 to 20 with a median value of 7 (Figure 1.9). Values of the Shannon's Information Index (H') calculated from family data ranged from 0.10 to 2.6 with a median value of 1.5. High values of H' were observed from sea level to 929m, always at plots with a substantial understory of green herbs, graminoids, or shrubs. Low values were observed at barren and recently snow-covered alpine sites.

Dominance of a single family was often quite high. The Berger-Parker index (d) calculated from family data ranged from 1.0 to 9.0 with a median value of 2.0. This means that a single family frequently represented at least half of the total number of specimens in a sample. As with Shannon's Information Index, low dominance (high evenness and high values of d) was observed from sea level to 929m at plots with a lush, green understory, and high dominance (low evenness and low values of d) were found at barren and recently snow-covered alpine sites.

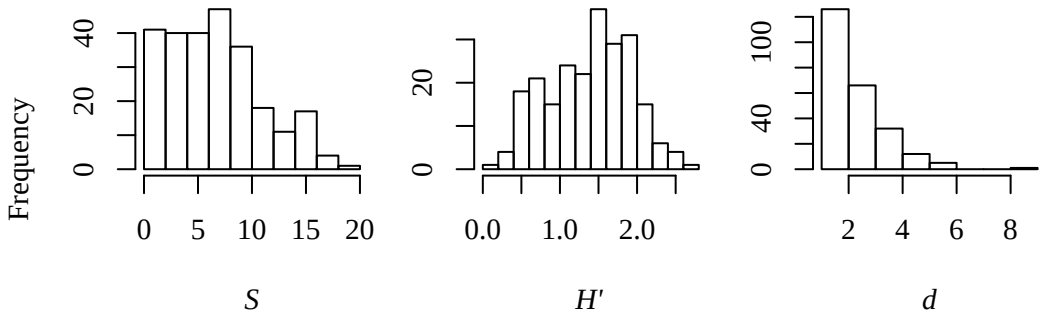

Figure 1.9: Histograms of diversity indices. family density (S), Shannon's Information Index (H'), and the Berger-Parker index (d).

Sites with high richness also tended to have high evenness (low dominance). Both the family density measure of richness and the Berger-Parker index measure of dominance were positively correlated with the composite index, Shannon's Information Index ($R = 0.73$ and $R = 0.80$, respectively). The Berger-Parker index was less strongly positively correlated with family density ($R = 0.42$). Along the progression from richness-dominated to dominance/evenness-dominated indices, correlation with the total number of specimens

(N_T) decreased (Table 1.5). Family density was most strongly correlated with the number of specimens ($R = 0.64$), and dominance was hardly correlated with the number of speicmens ($R = 0.07$).

Table 1.5: Correlation matrix of diversity indices and plot sums. N_T: total number of specimens collected per sample. N_A: number of specimens included in analyses. S: family density. H': Shannon's Information Index. d: Berger-Parker Index.

	$\boldsymbol{N_A}$	$\boldsymbol{S}$	$\boldsymbol{H'}$	$\boldsymbol{d}$
$\boldsymbol{N_T}$	0.94	0.64	0.14	0.07
$\boldsymbol{N_A}$		0.59	0.04	-0.14
$\boldsymbol{S}$			0.73	0.42
$\boldsymbol{H'}$				0.80

Nearby points tended to have similar values of richness and dominance/evenness, with similarity decreasing as the distance between points increased. All diversity indices showed strong spatial autocorrelation (Table 1.6); however, only a small portion of the variance was explained by spatial autocorrelation. About one third of the variance of the richness-dominated indices family density and Shannon's Information Index and about one fourth of the variance of the Berger-Parker index were explained by spatial autocorrelation. The distance at which spatial autocorrelation became negligible was about 60km for the Berger-Parker index and about 120km for family density and Shannon's Information Index (Figure 1.10).

Table 1.6: Spatial autocorrelation of diversity indices: results of inverse-distance-weighted Moran's I tests. S: family density. H': Shannon's Information Index. d: Berger-Parker index. Higher values of I indicate stronger spatial autocorrelation. $p > 0$ is the p-value of the test, the probability of obtaining the observed value of I under an assumption of no positive spatial autocorrelation.

Diversity Index	**Moran's $\boldsymbol{I}$**	$\boldsymbol{p > 0}$
$\boldsymbol{S}$	6.89	2.7×10^{-12}
$\boldsymbol{H'}$	4.70	1.63×10^{-5}
$\boldsymbol{d}$	3.18	2.00×10^{-6}

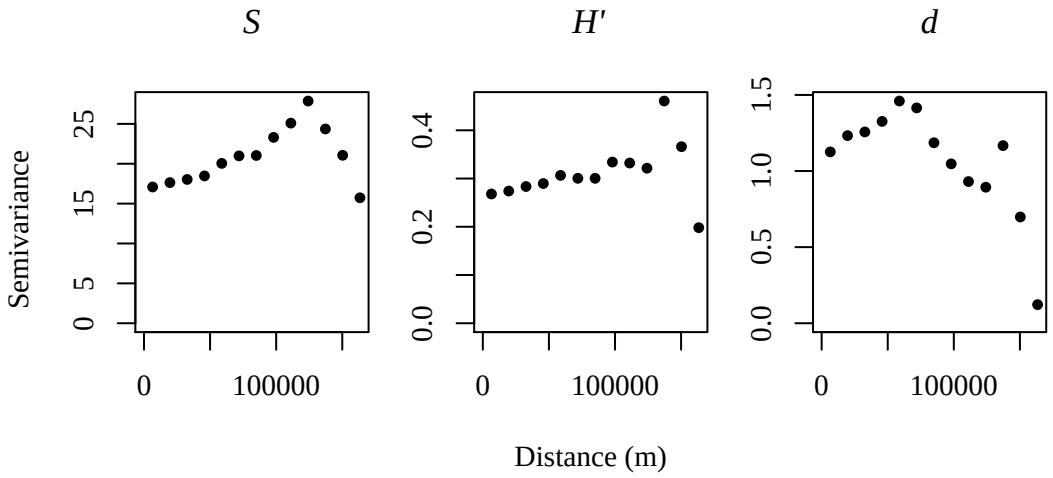

Figure 1.10: Spatial autocorrelation of diversity indices: empirical semivariograms. S: family density. H': Shannon's Information Index. d: Berger-Parker index. Semivariance is essentially a variance between observed values at pairs of points separated by some distance.

Random Forest Regression of Family Density

Random forest regressions incorporating only information that was available as continuous raster data explained only 22% of the variation of family density, but produced a reasonable map of predicted family density (Figures 1.11 and 1.12). The landcover classification of O'Brien (2006) was the most important variable, followed by NDVI (Figure 1.13). This raster predicted lowest arthropod family density at barren alpine sites. Highest densities were predicted near coastal wetlands at the north of the KENWR and productive mixed forests in the northwest of the KENWR.

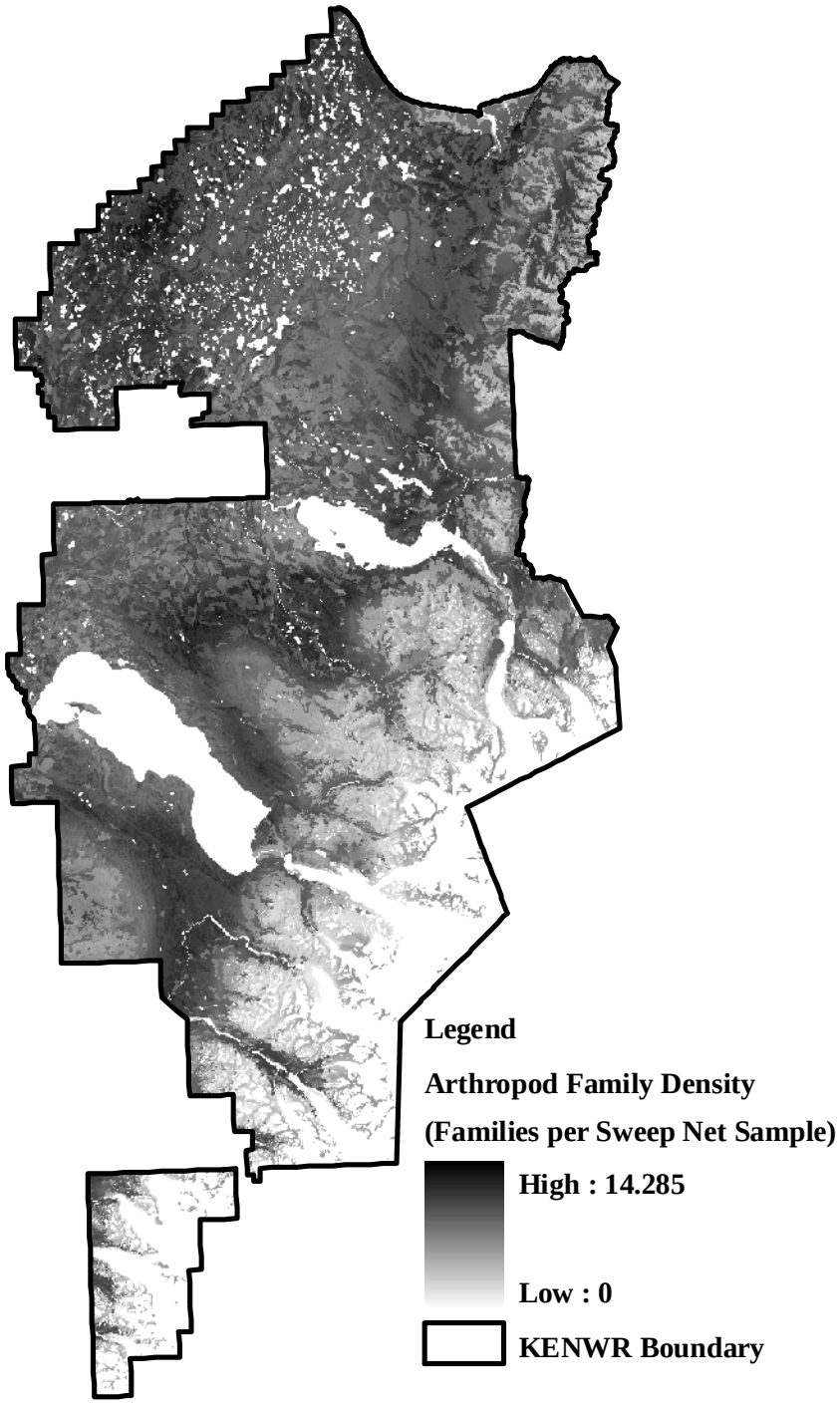

Figure 1.11: Arthropod family density predicted by random forest regression.

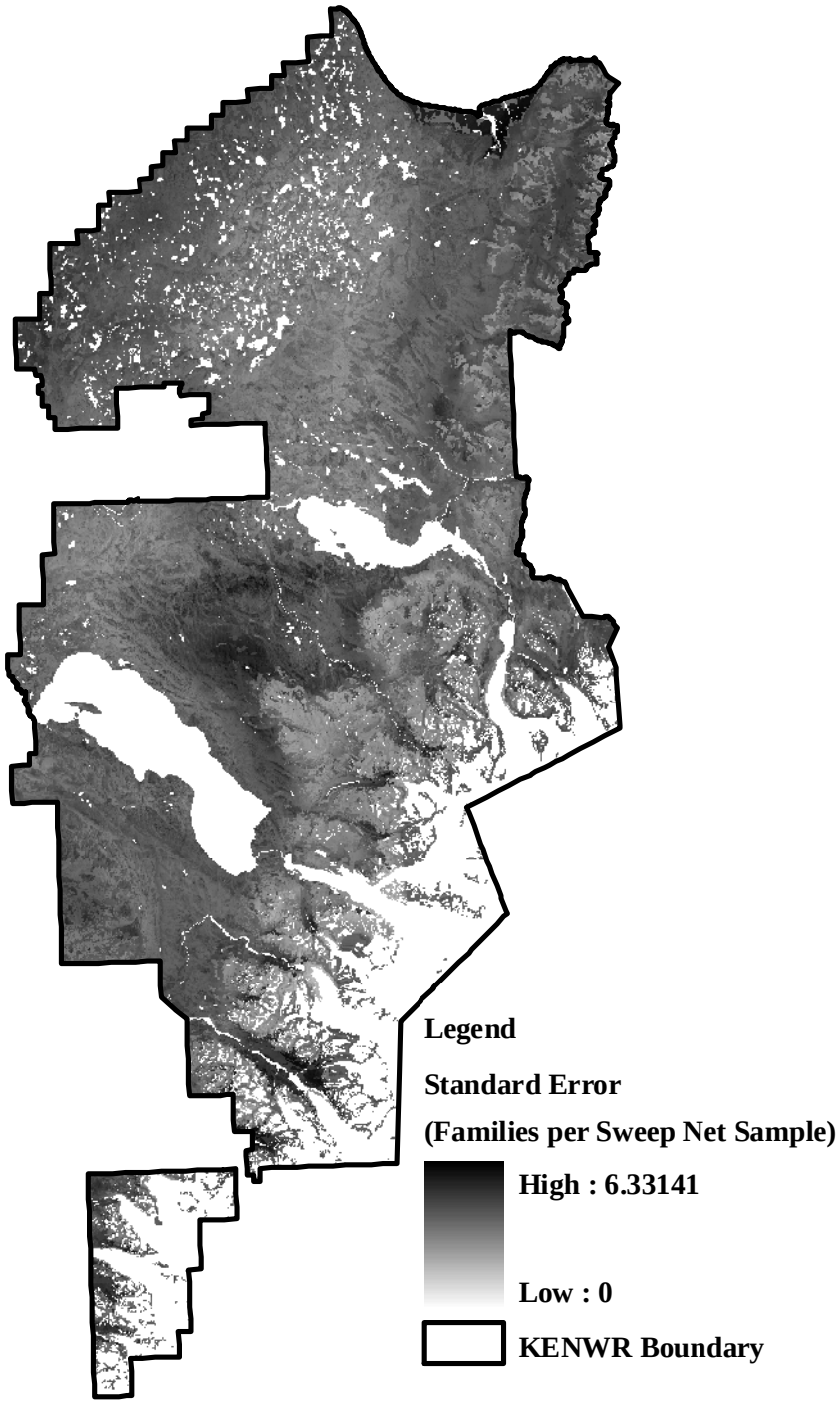

Figure 1.12: Standard error of arthropod family density predicted by random forest regression.

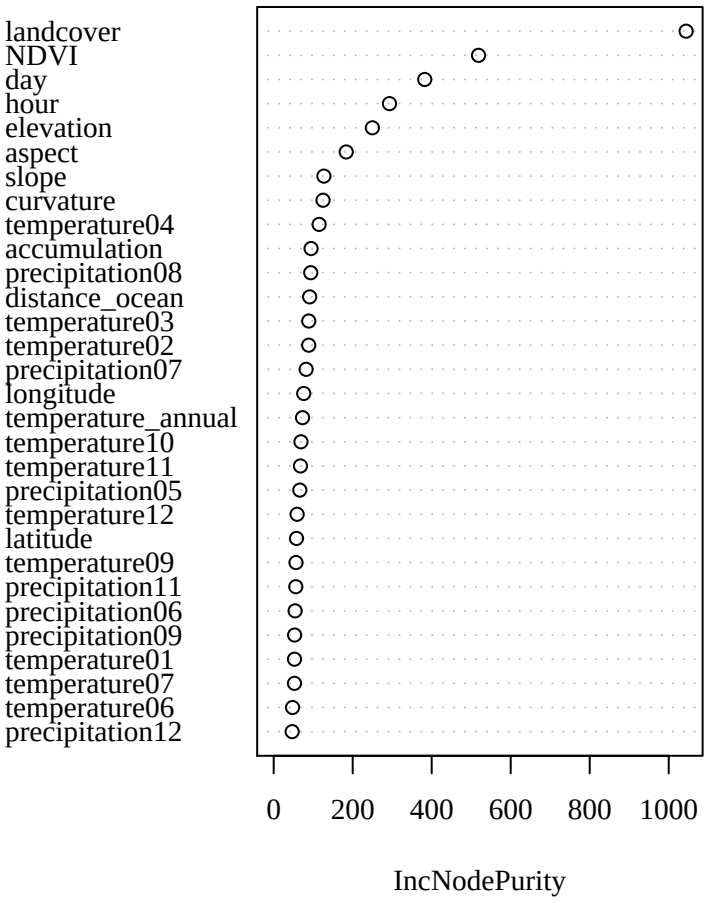

Figure 1.13: Importance of variables included in final random forest regression model. Variable names were defined in Table 1.3. IncNodePurity is the mean decrease in node purity, also known as the Gini splitting criterion (Breiman, 2002).

Arthropod family density was predicted to be highest (mean families/sweep net sample = 9.09) in the halophytic wetland habitat type (Table 1.7), followed by forest types with substantial proportions of hardwoods (except aspen) and shrubby wetlands. Diversity was predicted to be lowest at alpine, sparsely vegetated, and barren/rock habitat types (mean families/sweep net sample = 3.33–4.76). Note that no LTEMP points fell within the urban/cultural or aspen habitat types.

Table 1.7: Predicted arthropod family density by landcover type.

Landcover Type	Predicted Density (Families/Sweep Net Sample)	
	Mean	S.D.
Wetland - halophytic	9.03	0.90
Black cottonwood (balsam poplar)	8.88	1.17
Mixed forest	8.31	1.19
Paper birch	8.24	1.13
Wetland - shrub	8.15	1.44
Alder/Willow	8.05	1.54
Mixed deciduous	8.01	1.23
White/Lutz/Sitka spruce	7.99	1.33
Black spruce	7.79	1.09
Wetland - graminoid	7.76	1.24
Herbaceous	7.68	1.30
Urban/Cultural	7.54	1.34
Aspen	7.51	1.00
Willow	7.43	1.86
Mixed conifer	6.57	1.38
Alder	6.26	1.41
Other shrub	6.18	1.84
Mountain hemlock	6.06	1.16
Barren - wet	5.70	2.40
Alpine	4.76	1.26
Sparsely vegetated	3.69	1.48
Barren/Rock	3.33	0.85

Regression Analyses of Family Density

Bayesian Model Averaging Family density was positively related to S_{veg}, *Day*, P_{A}*T_{O}, and P_{A}**NDVI* and negatively related to P_{A}, h_{Alpine}, and h_{Hemlock} (Table A.1). Coefficients for these variables were included in at least 95.4% of models selected; and S_{veg}, *Day*, h_{Alpine}, and h_{Hemlock} were always included in selected models, indicating a high degree of certainty that family density was related to them. The coefficients of variation of these BMA estimates were fairly high, ranging from 0.16 for T_{O}:P_{A} to 0.37 for *NDVI*:P_{A}. A positive coefficient for the habitat type $h_{\text{Halophytic Wetland}}$ was also frequently included in models (49.7%).

Model Selection The five best models were consistent in including the terms S_{veg}, P_{A}*T_{O}, P_{A}**NDVI*, P_{A}, *Day*, h_{Alpine} and h_{Hemlock} (Table 1.8). The best model selected by the BMA

analysis was only slightly better than the next model, accounting for 6.0% and 5.6% of the posterior probability, respectively. Thus the two models performed comparably. They differed only in whether or not the positive coefficient $h_{\text{Halophytic Wetland}}$ was included.

Table 1.8: Best five negative binomial regression models of family density selected by Bayesian Model Averaging. nVar: number of variables. BIC: Bayesian Information Criterion. post prob: posterior probability.

Coefficient	model 1	model 2	model 3	model 4	model 5
(Intercept)	-1.3E+00	-1.4E+00	-1.4E+00	-1.2E+00	-1.5E+00
S_{veg}	1.9E-02	2.0E-02	2.1E-02	1.8E-02	2.2E-02
h_{alpine}	-5.7E-01	-5.7E-01	-5.3E-01	-6.4E-01	-5.3E-01
Day	1.8E-02	1.8E-02	1.8E-02	1.8E-02	1.8E-02
h_{hemlock}	-6.4E-01	-6.3E-01	-6.2E-01	-6.0E-01	-6.1E-01
P_{A}	-1.4E-03	-1.3E-03	-1.3E-03	-1.8E-03	-1.2E-03
T_{O}:P_{A}	9.7E-05	9.7E-05	9.2E-05	1.0E-04	9.1E-05
$NDVI$*P_{A}	1.1E-03	1.1E-03	1.7E-03	1.2E-03	1.7E-03
$h_{\text{halophytic wetland}}$	.	8.3E-01	.	.	7.9E-01
$NDVI$*E	.	.	-1.4E-03	.	-1.3E-03
E*P_{A}	.	.	.	3.6E-07	.
nVar	7	8	8	8	9
BIC	-1080	-1079	-1078	-1078	-1078
post prob	0.06	0.056	0.033	0.027	0.025

Best Model Selected In the best model selected by the BMA analysis ($R^2_{\text{dev}} = 0.51$, AIC = 1267.4, Table 1.9), observed family density was most strongly related to vegetation species density (S_{veg}), annual precipitation (P_{A}), observed plot temperature (T_{O}), NDVI ($NDVI$), Julian day (Day), and the habitat types alpine (h_{alpine}) and hemlock (h_{hemlock}).

Family density was negatively related to P_{A} and positively related to P_{A}*T_{O} and P_{A}*$NDVI$; these terms were clearly the most important explanatory variables, accounting for most of the deviance (35%) explained by the model. P_{A}*T_{O} was the most important of these terms. Plotting predicted values of S versus P_{A} at varying levels of T_{O} and $NDVI$ (Figures 1.14 and 1.15) showed that the relationship of S to P_{A} became more strongly positive as either T_{O} or $NDVI$ increased. In the best model, density was also positively related to Day, and negatively related to h_{Alpine} and h_{Hemlock}.

Table 1.9: Results of negative binomial regression model selected by BMA analysis. R^{2*}_{deva}: see definition on page 30. P_A: total annual precipitation. T_O: observed plot temperature. h_{alpine} and $h_{hemlock}$: the alpine and hemlock habitat types. S_{veg}: vegetation species density. Day: Julian day.

	Estimate	Std. Error	Pr(>|z|)	R^{2*}_{deva}
(Intercept)	-1.3E+00	8.8E-01	1.4E-01	0.00
P_A*T_O	9.7E-05	1.1E-05	< 2e-16	0.19
P_A	-1.4E-03	1.6E-04	< 2e-16	0.09
h_{alpine}	-5.7E-01	1.2E-01	3.0E-06	0.07
P_A*$NDVI$	1.1E-03	1.7E-04	6.7E-10	0.07
S_{veg}	1.9E-02	4.6E-03	4.9E-05	0.04
Day	1.8E-02	5.2E-03	5.7E-04	0.03
$h_{hemlock}$	-6.4E-01	1.7E-01	2.5E-04	0.02

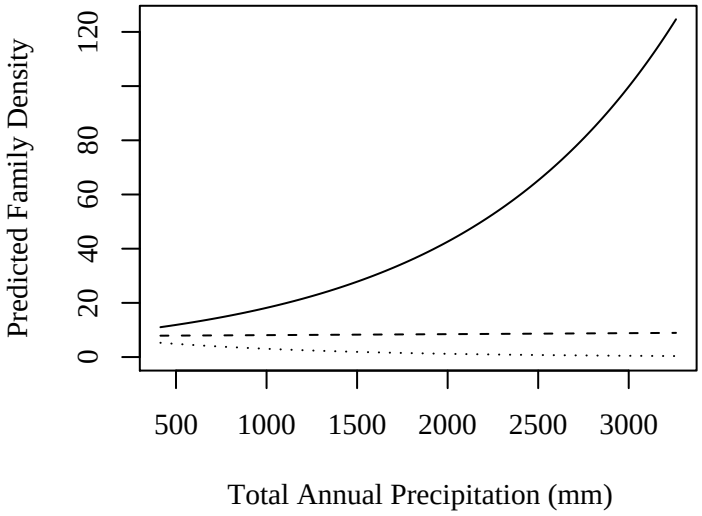

Figure 1.14: Predicted family density vs. total annual precipitation at varying values of observed plot temperature. The solid, dashed and dotted lines depict the predicted relationship at maximum (19.8°C), median (11.5°C), and minimum (1.5°C) values of observed plot temperatures, respectively. All other variables were set at median values.

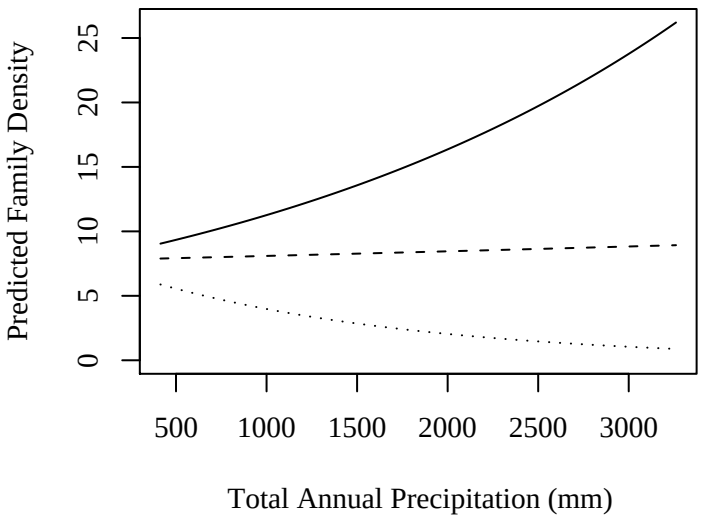

Figure 1.15: Predicted family density vs. total annual precipitation at varying values of NDVI. The solid, dashed and dotted lines depict the predicted relationship at maximum (0.57), median (0.26), and minimum (-0.40) values of NDVI, respectively. All other variables were set at median values.

Regression Analyses of Shannon's Information Index

Bayesian Model Averaging Shannon's Information Index, a composite measure of richness and evenness, was positively related to Julian day, time of day, vegetation species density, and NDVI (Table A.2). The relationship between Shannon's Information Index and observed plot temperatures was unclear based on the BMA analysis because of opposite signs on T_{O} and the interaction term $Hour{:}T_{\mathrm{O}}$. The terms Day, $NDVI{:}S_{\mathrm{veg}}$, $Hour{:}T_{\mathrm{O}}$, T_{O} and $Hour$ were included in at least 58.8% of models. All of these except for T_{O} and $Hour$ were included in at least 95% of models, indicating a high degree of certainty that relationships exist between Shannon's Information Index and these terms. The positive term $h_{\text{Halophytic Wetland}}$ was also included in 46.6% of models.

Model Selection The five best models of Shannon's Information Index all contained the terms Day, $NDVI{:}S_{\mathrm{veg}}$, and $Hour{:}T_{\mathrm{O}}$ (Table 1.10). Three of these five best models also included a positive $h_{\text{Halophytic Wetland}}$ term. The two best models performed similarly. The best model accounted for 10.5% of the posterior probability, not much more than the 8.6% accounted for by next best model.

Table 1.10: Best five Gaussian ordinary least squares regression models of Shannon's Information Index selected by Bayesian Model Averaging. nVar: number of variables. BIC: Bayesian Information Criterion. post prob: posterior probability. *Day*: Julian day. *Hour*: time of day (hours since midnight). T_{O}: observed plot temperature. *NDVI*: Normalized Difference Vegetation Index. $h_{\text{halophytic wetland}}$: halophytic wetland habitat type. *Sky*: sky condition.

Coefficient	model 1	model 2	model 3	model 4	model 5
Intercept	-3.5E+00	-3.5E+00	-5.0E+00	-5.0E+00	-4.2E+00
Day	3.6E-02	3.6E-02	3.5E-02	3.5E-02	3.1E-02
$NDVI{*}S_{\mathrm{veg}}$	4.5E-02	4.6E-02	4.9E-02	5.0E-02	4.8E-02
$Hour{*}T_{\mathrm{O}}$	2.2E-02	2.1E-02	3.1E-03	3.0E-03	2.9E-03
T_{O}	-1.6E-01	-1.6E-01	.	.	.
Hour	-2.0E-01	-2.0E-01	.	.	.
$h_{\text{halophytic wetland}}$	.	1.0E+00	.	1.0E+00	1.1E+00
Sky	.	.	.	.	-7.5E-02
nVar	5	6	3	4	5
r2	0.295	0.311	0.258	0.274	0.289
BIC	-52.01	-51.61	-51.10	-50.79	-49.89
post prob	0.105	0.086	0.067	0.057	0.036

Best Model Selected Temporal variables appeared to be the most important variables for explaining arthropod family diversity as measured by the composite diversity measure Shannon's Information Index, although it was also related to NDVI, observed plot temperatures, and vegetation species density. Shannon's Information Index was most strongly related to terms containing information about Julian day (*Day*), NDVI (*NDVI*), time of day (*Hour*), vegetation species density (S_{veg}) and observed plot temperature (T_{O}). The best model selected by the BMA analysis ($R^2 = 0.30$, AIC = 296.88) was composed of the positive terms *Day*, $NDVI{:}S_{\mathrm{veg}}$ and $Hour{:}T_{\mathrm{O}}$ and the negative terms *Hour* and T_{O} (Table 1.11). In this model, Julian day explained almost as much of the variation (14%) as the other terms combined (16%). The relationship between Shannon's Information Index, *Hour*, and T_{O} was not simple (Figure 1.16). Predicted values of Shannon's Information Index were highest for the combination of late morning census times and high plot temperatures but also fairly high for the combination of low plot temperatures and early census times. The model predicted low values of Shannon's Information Index on sites that were warm and visited early as well as on plots that were cool and visited in the afternoon, but very few data points fell within these regions of parameter space.

Table 1.11: Results of Gaussian oridanary least squares regression model selected by BMA analysis. R_a^{2*}: see definition on page 30. *Day*: Julian day. *Hour*: time of day (hours since midnight). T_{O}: observed plot temperature. *NDVI*: Normalized Difference Vegetation Index.

	Estimate	SE	Pr(>|z|)	R^2
(Intercept)	-3.5E+00	1.1E+00	-3.2E+00	0
Day	3.6E-02	5.5E-03	6.5E+00	0.14
Hour	-2.0E-01	7.0E-02	-2.8E+00	0.03
T_{O}	-1.6E-01	4.7E-02	-3.4E+00	0.00
$Hour{*}T_{\mathrm{O}}$	2.2E-02	5.8E-03	3.8E+00	0.06
$NDVI{*}S_{\mathrm{veg}}$	4.5E-02	9.8E-03	4.6E+00	0.07

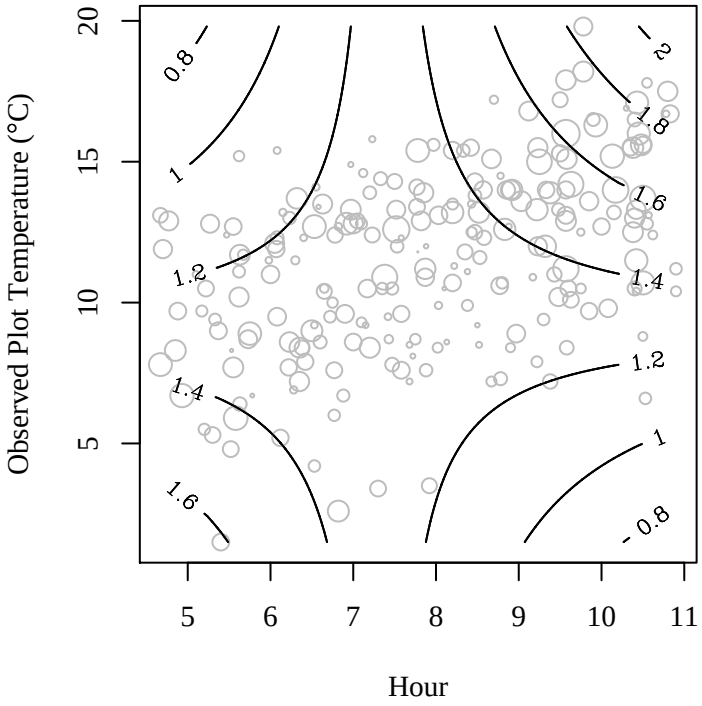

Figure 1.16: Predicted values of Shannon's Information Index over the range of values of observed plot temperature and time of day. The surface of predictions are output of the best model given median values of other variables. Grey circles represent observed combinations of plot temperatures and time of day. The diameters of the circles are proportional to the observed values of Shannon's Information Index ($H'_{\mathrm{max}} = 2.6$ and $H'_{\mathrm{min}} = 0.1$).

Regression Analyses of the Berger-Parker Index

Bayesian Model Averaging Evenness (here measured as the inverse of dominance) was consistently positively related to Julian day, the interaction of NDVI and vegetation species

Table 1.12: Best five gamma regression models of the Berger-Parker Index selected by Bayesian Model Averaging. nVar: number of variables. BIC: Bayesian Information Criterion. post prob: posterior probability. *Day*: Julian day. *Hour*: time of day (hours since midnight). T_{O}: observed plot temperature. *NDVI*: Normalized Difference Vegetation Index. h_{Snow}: snow wetland habitat type. S_{veg}: vegetation species density. *Sky*: sky condition. *E*: Elevation

Coefficient	model 1	model 2	model 3	model 4	model 5
Intercept	2.4E+00	2.5E+00	2.0E+00	2.1E+00	2.1E+00
Day	-1.1E-02	-1.1E-02	-9.1E-03	-9.5E-03	-9.2E-03
$NDVI{*}S_{\mathrm{veg}}$	-1.5E-02	-1.4E-02	-1.3E-02	-1.3E-02	-1.4E-02
$T_{\mathrm{O}}{*}E$	.	.	-2.9E-05	-2.9E-05	-3.0E-05
$T_{\mathrm{O}}{*}Hour$	-1.1E-03	-1.0E-03	.	.	.
h_{Snow}	.	5.3E-01	5.6E-01	5.5E-01	.
$Day{*}E$	.	.	.	1.7E-06	1.8E-06
E	.	.	2.8E-04	.	.
nVar	3	4	5	5	4
BIC	-1054.00	-1054.00	-1054.00	-1054.00	-1054.00
post prob	0.026	0.021	0.021	0.021	0.02

density, and the interaction of observed plot temperatures and elevation (Table A.3). The terms *Day* and $NDVI{*}S_{\mathrm{veg}}$ were included in at least 82.9% of selected models, indicating a high degree of certainty that evenness was related to them. $T_{\mathrm{O}}{*}E$ and $T_{\mathrm{O}}{*}Hour$ were also frequently included (61.9% and 48.8%, respectively). The coefficient of variation of the BMA estimates was quite high, ranging from 0.40 for *Day* to 1.4 for $T_{\mathrm{O}}{*}Hour$.

Model Selection The best model was selected with a low degree of confidence. Several models performed similarly, each model representing a very small portion of the posterior probability (Table 1.12). The five best models each accounted for only 2.0–2.6% of the posterior probability. These five models consistently included the terms *Day*, $NDVI{*}S_{\mathrm{veg}}$, and some interaction term involving T_{O}.

Best Model Selected Evenness was positively related to Julian day and the interaction terms $Hour{*}T_{\mathrm{O}}$ and $NDVI{*}S_{\mathrm{veg}}$ in the best model selected ($R^2_{\mathrm{dev}} = 0.20$, AIC = 617.6, Table 1.13). As in models selected for Shannon's Information Index, Julian day was the most important variable in terms of the amount of deviance explained (10% in the best model). This model would predict greatest evenness (lowest dominance) when samples are

taken late in June rather than early in June, later in the morning rather than early, when temperatures are warm, and where productivity and vegetation species density are high.

Table 1.13: Results of gamma regression model selected by BMA analysis. Note that the signs of coefficients are reversed in inverse gamma regressions so that a negative sign indicates a positive relationship. R^{2*}_{deva}: see definition on page 30. *Day*: Julian day. T_{O}: observed plot temperatures. *Hour*: time of day (hours since midnight). *NDVI*: Normalized Difference Vegetation Index. S_{veg}: vegetation species density.

	Estimate	SE	Pr(>|z|)	R^2_{deva}
(Intercept)	2.4E+00	3.6E-01	6.8E+00	0.00
Day	-1.1E-02	2.1E-03	-5.1E+00	0.10
T_{O}*$Hour$	-1.1E-03	2.9E-04	-3.7E+00	0.04
$NDVI$*S_{veg}	-1.5E-02	3.3E-03	-4.6E+00	0.07

Comparison of Diversity Indices

Consistent Patterns Observed plot temperature, Julian day, NDVI, and vegetation species density were positively related to all three diversity indices and were consistently included in best models (Table 1.14). Through the progression from a completely richness-dominated measure of diversity (S) to a measure of only dominance/evenness (d), the relative importance of annual precipitation decreased and that of temporal variables increased. The importance of temporal variables corresponded with the degree to which abundance contributed to calculation of diversity indices.

Note that the diversity indices most influenced by temporal variables were those in which abundance was taken into account.

Table 1.14: Comparison of variables in regression coefficients most often chosen by BMA analyses. Signs (+ and −) indicate signs of coefficients. The labels "Richness" and "Evenness" categorize the diversity indices based on the aspects of diversity they reflect. Asterisks (∗) are used to point out variables that were included in the best models selected by BMA analyses for all three diversity indices. S: family denisty. H': Shannon's Information Index. d: Berger-Parker index. Day: Julian Day. S_{veg}: vegetation species density. $NDVI$: Normalized Difference Vegetation Index. $Hour$: time of day (hours since midnight). T_{O}: observed plot temperatures. P_{A}: total annual precipitation (PRISM climate model). h_{alpine} and h_{helmock}: the alpine and hemlock habitat types.

	Richness	Evenness	
Variable	S	H'	d
Day	$+^*$	$+^*$	$+^*$
S_{veg}	$+^*$		
$NDVI$*S_{veg}		$+^*$	$+^*$
$NDVI$*P_{A}	$+^*$		
T_{O}*P_{A}	$+^*$		
T_{O}*$Hour$		$+^*$	$+^*$
$Hour$		$-$	
T_{O}		$-$	
P_{A}	$-$		
h_{alpine}	$-$		
h_{hemlock}	$-$		

Annual Precipitation Most of the points where extremely low diversity was observed received high ($>$ 500m) precipitation (Figure 1.17). The diversity indices H' and d could not be calculated for points where no arthropods were collected ($S = 0$). In addition, if only a single family was collected on a plot, then $S = 1$, $d = 1$, and H' could not be calculated. For this reason, regressions of H' and d did not contain much information about high precipitation sites, but regressions of S included the entire range of precipitation values observed.

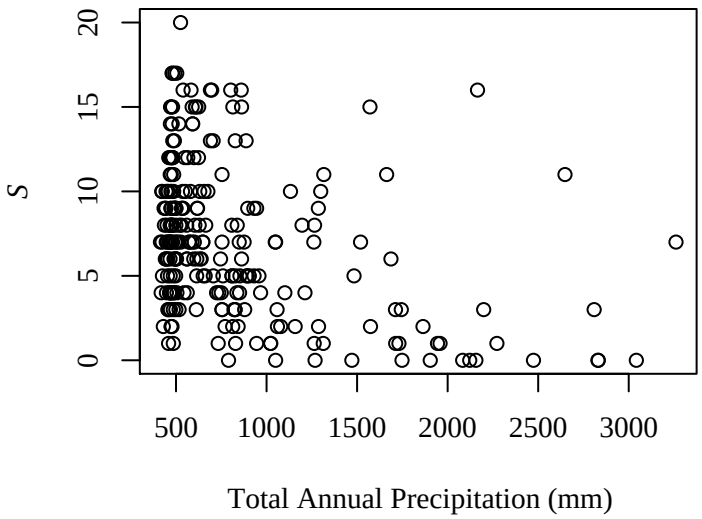

Figure 1.17: Arthropod family density versus total annual precipitation.

The Hemlock Forest and Alpine Habitat Types Coefficients for these habitat types were only included in models of S, but they were important terms in these models, accounting for a substantial portion of the deviance (9%) explained by the best model. This indicated that arthropod family density was lower for these habitat types than would have been expected based on values of other variables.

The Halophytic Wetland Habitat Type Although this habitat type was not included in any of the best models selected, it was included in nearly half of all models of S and H'. The relationship of diversity indices to this habitat type was always positive, indicating that diversity in salt marsh habitat types was greater than would have been expected based on other variables. This habitat type was represented by a single data point.

Regressions of Observed Temperatures and NDVI

Observed plot temperatures were positively related to the time of day and negatively related to elevation. The best model selected by a BMA analysis of observed temperatures was a simple model including the main effects E and $Hour$ and the interaction term E^*Day ($R^2 = 0.36$, Table 1.15). A positive relationship with $Hour$ was most important in terms

of the amount of variation explained ($R^{2*} = 0.18$). Note that T_A was not included in this model.

Table 1.15: Results of regression of observed temperatures.

	Estimate	SE	Pr(>|z|)	R^{2*}
(Intercept)	5.154	8.689e-01	1.00e-08	0
E	-6.353e-02	1.171e-02	1.35e-07	0.11
$Hour$	9.336e-01	1.079e-01	6.38e-16	0.18
$E*Day$	3.504e-04	6.861e-05	6.52e-07	0.07

The NDVI measure of productivity was negatively related to elevation and annual precipitiation, but since elevation and annual precipitation were strongly correlated ($R = 0.83$), results of multivariate regressions were unstable. Coefficients for elevation and annual precipitation were the most important variables in BMA analyses in terms of the amount of variance they explained and the frequency with which they were included in models, but their values and signs varied widely depending on what other variables were included and whether or not an intercept term was included. Simple regressions of $NDVI$ versus E ($\hat{\beta}$ = -2.75×10^{-4}, $p < 0.001$, $R^2 = 0.28$) and $NDVI$ versus P_A ($\hat{\beta}$ = -2.47×10^{-4}, $p < 0.001$, $R^2 = 0.41$) were much easier to interpret, each with negative coefficients explaining a substantial portion of the variation in NDVI.

Spatial Autocorrelation of Residuals

Spatial autocorrelation of residuals increased along the gradient from richness-dominated to evenness-dominated diversity indices. The selected regression models of arthropod family diversity explained all of the spatial autocorrelation of family density (S), but only some of the spatial dependence of evenness (d). Residuals of the best model of arthropod family density were not spatially correlated, residuals of the best model of Shannon's Information Index were not significantly positively correlated (though apparently slightly more spatially correlated than residuals of the best model of S), and residuals of the best model of the Berger-Parker index were positively spatially correlated (Table 1.16). Empirical variograms of the residuals of family density and Shannon's Information Index were flat at distances less than about 100km (Figure 1.18), indicating no spatial autocorrelation at distances of at least the 4.8km spacing between sampling points. The empirical variogram of the residuals from

the model of the Berger-Parker index indicates a small amount of spatial autocorrelation (up to about one sixth of the overall variance), becoming negligible at distances of about 40km.

Table 1.16: Spatial autocorrelation of residuals of diversity models: results of inverse-distance-weighted Moran's I tests. S: family density. H': Shannon's Information Index. d: Berger-Parker index. Higher values of I indicate stronger spatial autocorrelation. autocorrelation. $p > 0$ is the p-value of the test, the probability of obtaining the observed value of I under an assumption of no positive spatial autocorrelation.

Diversity Index	**Moran's I**	$p > 0$
S	-1.12	0.868
H'	1.38	0.084
d	2.13	0.017

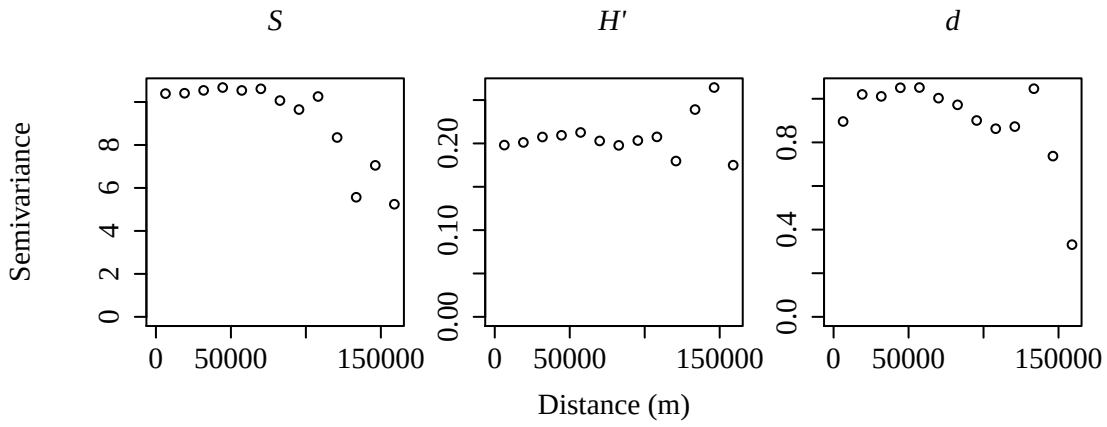

Figure 1.18: Spatial autocorrelation of residuals of diversity models: empirical semivariograms. S: family density. H': Shannon's Information Index. d: Berger-Parker index. Semivariance is essentially a variance between observed values at pairs of points separated by some distance.

1.4 Discussion

Overview of the Fauna Collected

The composition of the fauna collected in this study was typical of North American boreal forest and arctic tundra biomes and consistent with the arthropod fauna of Canada (Danks, 1979, 1988; Danks and Foottit, 1989) in that Diptera dominated. The biting flies, especially

mosquitoes (Culicidae) were disproportionately super-abundant in the samples because they were attracted to the collectors. Members of most other groups avoided being collected to some degree or were stationary. Particularly agile and alert arthropods, such as dragonflies, adult butterflies, and many flower flies (Syrphidae) evaded capture and were certainly under-represented.

Hemiptera, especially aphids (Aphididae) were also quite abundant. This is consistent with Werner (1983), who found that Hemiptera often represented a high proportion of arthropod biomass relative to most other orders in boreal forests of interior Alaska. Perhaps most noteworthy was the abundance of globular springtails (Collembola: Sminthuridae). I have not found mention of the Sminthuridae being an especially abundant group on vegetation in our region.

It was clear that these samples best represent the arthropod fauna that tend to be present on the surface of vegetation. For example, some common groups of beetles that live mainly in litter (e.g., Carabidae), aquatic habitats (e.g., Dytiscidae and Hydrophilidae), or in wood (e.g., Scolytinae and Cerambycidae) were seldom or never collected, while Cantharids, which hunt on leaves of shrubs and herbs, were frequently collected.

Overview of Diversity Patterns

Clear patterns were discerned in the distribution of arthropod family diversity over the KENWR. Observed athropod diversity at the scale of a $100m^2$ plot was influenced by local climate, productivity, and time of sampling.

Seasonal Phenology, Daily Activity Patterns, and Temperature

As predicted, indices of observed arthropod diversity were strongly influenced by seasonal phenology even in my three week June sampling window (June 7–30). That arthropod diversity was increasing over this period suggests arthropod diversity had not passed its seasonal peak. Note that, since no terrestrial arthropod species in this region is known to be migratory, the observed seasonal phenology pattern was entirely an issue of detection probability changing over time. The seasonal life history and activity patterns of arthropods were such that the likelihood of collecting them was generally increasing over this sampling window.

I also found strong evidence for increasing arthropod activity of the daily sampling period (from 04:40 to 10:54 hours). Most arthropods were apparently less active in the dusky, cool early morning hours than in the warmer, sunny mid-morning hours. The observed pattern is likely predominantly driven by increasing detection probabilities over the course of the morning rather than actual changes in diversity driven by dynamics of immigration and emigration. However, a small number of strong-flying groups, such as bumble bees, might have shown daily patterns of immigration and emigration through their foraging efforts.

There are two explanations for why observed plot temperature (T_{O}) was often included in models while average annual temperature (T_{A}) was rarely included. The first is an issue of temporal scale: arthropods appeared to respond to changes in temperatures over the course of a day. Plot temperatures observed at the time of sampling captured this among-day and within-day variation, while average annual temperatures provided an estimated summary of climate over the long term. Resident arthropods at the site likely responded to increasing temperatures over the course of the day by increasing their activity, which may have increased detection probabilities. The second reason is an issue of how well the two different measurements of temperature represented the small plots: observed plot temperatures were taken on the plots, so the spatial scale and spatial alignment matched the plot size; estimates of average annual temperature were generalized over coarse 2 km grid cells in which there may have been substantial variation in climate, especially in mountainous terrain. In addition, predictions of average annual temperatures from the PRISM model may have been inaccurate representations of local climate, especially considering that the PRISM model predictions used here were made using data from a handful of climate stations in the vicinity of the KENWR (see maps of input data for the PRISM model in Simpson et al., 2005).

The indices that measured evenness, which took into account relative abundance, were more influenced by temporal variables than indices that measured richness, which only considered presence or absence of taxa. It appeared that arthropods became more active and could be more easily collected as local temperatures warmed (i.e., detection probabilities generally increased with the hour of the day and with increasing temperature). Since detection probabilities could not be estimated from this dataset, it was not possible to distinguish between increasing diversity or increasing detection probabilities with increasing observed

temperatures. The correlation of temporal variables and observed temperatures as well as the inclusion of temporal variables in models support the latter explanation, but it is reasonable to infer that both are true.

Precipitation

A relationship with precipitation was observed for only one of the three diversity indices, S, indicating that precipitation was related to richness, but not dominance/evenness. At least part of the reason that annual precipitation was not included in models of H' and d was the issue of calculability. Arthropod family density was often very low at these high elevation, high precipitation sites so that observed S was 0 or 1. In this way, extremely high precipitation seemed to act more as a predictor of whether or not arthropods would be collected at all than a measure of diversity (although 0 and 1 are valid values of density).

The observed complicated relationships among family density, annual precipitation, observed plot temperatures, and NDVI can be interpreted in light of the relationships of climate and elevation on the KENWR. Observed temperatures and NDVI were generally lowest at high elevations, where precipitation was highest. At these elevations, the observed relationship between family density and annual precipitation was negative. Since most of the precipitation at these high sites falls as snow, this may be interpreted as a negative relationship between arthropod richness and increasing snow loads (climatic severity). When observed temperatures or NDVI was high, the observed positive relationship of family density with annual precipitation agreed with my prediction. Places that were both wet and warm or wet and productive had higher diversity of arthropods than places that were either dry and warm or dry and less green.

Elevation

I found little support for a simple relationship between arthropod diversity and elevation in full models of any of the three diversity indices. Although no direct relationships with elevation were found, climatic variables and NDVI were correlated with elevation so that diversity was related to elevation indirectly. This could be interpreted as climate and productivity being the more proximal causes of the observed patterns of diversity while elevation was important in determining climate and resulting patterns of productivity.

Productivity

There are several plausible reasons why productivity was a strong predictor of arthropod diversity: bias of the sweep net sampling method, relationships of NDVI to other variables, the high resolution of NDVI data, and real relationships between arthropod diversity and vegetative productivity. Sweep net samples were clearly biased, yielding more arthropods where there was plenty of green vegetation to sweep than where there was very little vegetation. In this way, sweep net samples were more representative of the community of arthropods on vegetation except tree canopies rather than a good representation of the total fauna of these sites. NDVI measurements reflect a host of biologically relevant variables not directly measurable by other means. Any factor that affects plant growth, such as soil type, drainage, water availability, snow cover, fire history, and local climate, may be reflected in NDVI observations, so it is possible that observed arthropod values of diversity were related to one or more of these variables for which NDVI was a surrogate rather than productivity causing the observed arthropod diversity pattern. However, there was good reason to expect high productivity to lead to high arthropod diversity based on the hypothesis of Wright (1983). The hypotheses which predicted unimodal relationships of diversity to productivity would predict simple positive relationships if only the increase phase of the curve was examined. Since maximum observed values of NDVI in this dataset were only moderately high (based on the scale of values listed by Myneni et al., 1997), this possibility cannot be ruled out. There is in this dataset at least support for the existence of an overall positive relationship between arthropod diversity and productivity over the landscape of the KENWR.

Based on the current analysis, it is difficult to distinguish between the competing diversity-productivity hypotheses that predict monotonically increasing or unimodal patterns. My results were consistent with the species-energy theory of Wright (1983). Investigation of the species-productivity-area hypothesis of Rosenzweig (1995) would require examination of how the size of patches changed with the productivity of patches, which was not attempted here. The unusually low arthropod diversity of the structurally simple (at least from 0-2m above the ground) hemlock and alpine habitat types and the explanatory power of vegetation species density support the habitat heterogeneity hypothesis of Tilman (1982). However, the high arthropod diversity of the structurally simple and vegetatively monotonous halophytic wetland habitat type conflicts with this explanation. Other

productivity-diversity hypotheses that were based on competition, predation, and disturbance (e.g., Leibold, 1996; Huston, 1979; Rosenzweig, 1995) could not be addressed with the above analyses.

Vascular Plant Species Density

I predicted that arthropod diversity would increase with vascular plant species density because greater species diversity of vascular plants would provide more kinds of niches to be occupied. I found that all indices of arthropod diversity increased with vascular plant species density in both BMA analyses and single-variable regressions, supporting this hypothesis and consistent with the results of Murdoch et al. (1972). However, it was not possible to distinguish between this hypothesis and another explanation: arthropod family diversity and vascular plant species density may be determined by the same underlying processes, such as the aforementioned diversity-productivity and diversity-climate relationships.

The Hemlock Forest Habitat Type

While NDVI measurements were high for hemlock forests, measured arthropod diversity was relatively low. In hemlock forests, the understory where sweep-net samples were collected had sparse and uniform vegetation with low vegetation species diversity. Fewer than the expected number of arthropod families were collected because vegetation density and vegetation species diversity in the understory were low despite high observed values of NDVI.

The Alpine Habitat Type

There are three reasons why arthropod family richness was low in alpine habitats. The seasonal phenology of vegetation in alpine sites was often far behind that of lower sites. The vegetation had not yet greened up when many of the alpine sites were sampled in June. Resident arthropods at these sites may have had similarly delayed phenologies. Mobile arthropods, such as pollinators, would not visit these alpine sites until flowering occurred. The low height and sparseness of vegetation on many alpine sites may also have reduced the efficacy of sweep net sampling, with most alpine arthropods concentrated close the substrate. The third possibility is that diversity was low in the alpine, probably due to

climatic harshness of this biome and possible because of the relative structural simplicity of this habitat type, which had few woody trees or shrubs.

The Halophytic Wetland Habitat Type

Arthropod diversity was greater in a salt marsh site than would have been expected based on other variables. Although coastal salt marshes are not particularly productive in terms of NDVI and the graminoid-dominated vegetative community is not diverse in terms of species or structure, these habitats may support a diverse communities of arthropods, mainly Diptera (Cheng, 1976). Part of the reason for the high observed diversity of this habitat type had to do with the relative efficiency of the sweep net sampling method. The salt marsh was dominated by Diptera, which appeared to be sampled quite well by sweep net.

Spatial Autocorrelation

Essentially all of the spatial autocorrelation in arthropod family richness was accounted for by explanatory variables. However, not all of the spatial autocorrelation of dominance/evenness was explained by these variables. This suggests that some part of the variation in dominance/evenness of arthropods is related to a spatially autocorrelated variable that was not considered in this analysis.

It should be noted that the methods used here tend to underestimate spatial autocorrelation and, consequentially, overstate precision of regression parameter estimates. An assumption of the regression models was that observations were independent. Spatial autocorrelation reduces independence of observations, effectively reducing sample size and precision of estimates. Such correlation of observations in regression models that assume independence produces underestimates of standard errors of regression coefficients and confidence intervals of predictions that are too narrow. Compared to the two-step approach used here, model fitting methods that simultaneously estimate regression coefficients and spatial parameters may result in larger confidence intervals for regression coefficients and predictions, and they will often indicate a stronger influence of spatial autocorrelation.

Prediction by Random Forest Regression

Where plot data, such as observed temperatures, were not available, the best predictor of arthropod family density was habitat type as summarized by the vegetation classification of O'Brien (2006). Arthropod family density was predicted to be highest in the halophytic wetland habitat type, a habitat in which high family density was observed. This is somewhat surprising considering that this habitat type had only moderate productivity (as measured by NDVI), low vegetation species density, and little variation in vegetation structure. High diversity was also predicted in habitat types dominated by deciduous hardwoods and shrubs, but it is unclear whether this was due to the generally high productivity of these habitat types, high vegetation species richness, or differences in structural complexity. Lowest arthropod diversity was predicted in barren, sparsely vegetated habitats, consistent with the low productivity, relatively low vegetation species density, and simple structure of these habitats.

Synthesis

It is generally accepted that the observed trends of diversity over latitudinal and elevational gradients are not the results of latitude or elevation themselves, but rather of climate regimes influenced by latitude and elevation. The broad spatial extent and large number of data points where diversity data were collected as well as the availability of climate data over the landscape allowed me to directly examine the relationships between diversity and climate variables rather than looking at relationships with elevation as a surrogate for climatic gradients. The results suggest that diversity is highest at the most favorable sites, characterized by warm temperatures, high productivity, and high plant species diversity. Such sites may support larger, more stable populations that are less prone to local extinction. This in turn would allow diversity to accumulate in these sites. The most favorable sites seemed to be defined not by any single variable, but by combinations of favorable conditions.

These findings are consistent with predictions of most of the diversity-latitude, diversity-elevation, diversity-productivity, or arthropod diversity-vegetation diversity hypotheses; they highlight the importance of the more proximal influences of local climate, productivity and vegetative diversity and especially the importance of their interactions. They best supported

the temperature and water availability hypothesis of McCain (2007) and the species-energy theory of (Wright, 1983).

My results are consistent with the conclusions of Parmesan and Yohe (2003), who found that climate variables were important for determining arthropod distributions. On the KENWR, regional climate has become warmer and drier in the latter part of the 20^{th} century compared to long-term averages (Berg, 2000). Since arthropod richness was positively related to temperature and generally negatively related to precipitation, it is difficult to predict how arthropod diversity will respond to a generally warmer, but drier climate. Due to the climatic constraints on the distributions of at least some species such as Carabid beetles (Ashworth, 2001), these species would be expected to generally move up slope as the climate warms. The expected warming and drying trend would also be expected to push the elevation of maximum diversity up slope based on the diversity-climate model of McCain (2007).

The effects of climate on individual species makes the influence of climate warming on arthropod diversity difficult to forecast. Warming of the cold climate of the KENWR would make this region more susceptible to successful invasion of exotic species from more temperate regions. Although this may initially increase diversity, invasive species (including non-arthropod taxa and diseases) could potentially disrupt native communities to the extent that native species would decline, decreasing diversity. Under changing climate, even native species might alter their behavior enough to dramatically transform the character and behavior of large areas. For example, warming temperatures over recent decades have enabled the native spruce bark beetle to kill essentially all mature white spruce trees over vast areas of the KENWR (Berg, 2000). This in turn has changed regions of mature coniferous forest, where fires were extremely rare, into grass-dominated woodlands characterized by frequent, landscape-scale fires. Although the behavior of this relatively well-studied spruce beetle might be somewhat predictable under expected climate regimes, hardly anything is known about the biology of most arthropod species of the region, so it is not possible to predict how most of arthropod diversity on the KENWR will respond to climate change.

Concerns/Caveats

In terms of geographic extent, scale of plots, taxonomic breadth, and taxonomic resolution, the methods of this present study differ from most other studies of arthropod diversity which attempt to find relationships between arthropod diversity and environmental explanatory variables. The most common methods involve narrower taxonomic breadth (often one or two orders), higher taxonomic resolution (usually species), relatively large plots, few sites, and a small area of inference. Bailey et al. (2004) appears to be the closest analogue of the present study. They sampled butterfly species richness at 195 large sites over the extent of a basin. Mac Nally et al. (2003) was another comparable study in which 49 relatively large sites over the extent of a landscape were inventoried for Lepidoptera. In the present study, taxonomic resolution and completeness of inventories at each site were sacrificed for a higher number of sites at relatively low cost. The small plots corresponded with individual pixels of raster grids, unlike Bailey et al. (2004) and Mac Nally et al. (2003), in which sites covered many pixels so that summary statistics were calculated for each site. In this way the design worked at a finer spatial resolution at a broad scale.

I desired to make inferences for as much of the arthropod fauna as possible, so the taxonomic scope chosen was as broad as was feasible given available sampling methods, literature and expertise. The sweep net sampling method was chosen because it could be executed quickly and yielded a broad selection of arthropods. Still, since no single method samples all groups well, taxonomic scope was limited by this choice. Many groups of arthropods that do not normally inhabit standing vegetation (e.g., soil arthropods, aquatic arthropods, and parasites of vertebrates) were rarely or never collected. Additional groups were excluded for practical reasons described in the methods section.

Aside from complete inventories where a wide variety of methods would be employed at each site over the duration of the growing season, accurate measurements of arthropod diversity are not possible. Such complete inventories were not feasible over the LTEMP grid. I chose methods to make the taxonomic scope as wide as possible given practical and logistic constraints so that generalizations could be made for much of the arthropod fauna of the KENWR. In doing so, taxonomic scope of this study was broader than in most studies of arthropod diversity. This appears to be the only study of such broad scope over the extent of a landscape.

The coarse, family level resolution of this study was less than ideal. Higher taxonomic resolution would have provided more information, possibly allowing more and stronger statistical relationships to be discerned. In cases where varying taxonomic resolutions have been compared (e.g., Balmford et al., 1996; Bowman and Bailey, 1997; Marshall et al., 2006), taxonomic resolution generally did not alter conclusions. Since the cost of identifications increases dramatically with finer taxonimic resolutions (Balmford et al., 1996; Marshall et al., 2006), I chose a practical level of resolution.

In some ways a compromise was made between resolution and scope because many specimens not identifiable at the species level (e.g., damaged and immature specimens) were easily sorted to the family level. In this respect taxonomic breadth was increased at the cost of lower resolution.

The sweep net sampling method was chosen because it sampled a broader taxonomic scope than most other methods considered and it could be executed quickly. How this affected taxonomic scope was discussed above. Another consequence was that conclusions over the landscape, such as the observed relationships between arthropod diversity and NDVI, should be interpreted in the light of potential biases. Since sweep net samples were collected directly from the vegetation, it would be expected that more arthropods would be collected from lush, thickly vegetated habitats than from sparse, barren habitats simply due to differences in the amount of substrate sampled. In places where most arthropods would be at ground level (e.g., barren gravel) or in tree canopies beyond the reach of nets, sweep net samples were probably less representative of the arthropod community than at sites dominated by thick, low vegetation.

No single method could have avoided such biases. The only way to gain a fairly complete inventory at each site would have been to employ multiple complimentary methods (e.g., malaise traps, pitfall traps, UV light traps, and Berlese funnels) over the course of the field season at each site, which would have allowed only a handful of sites (less than ten) to be sampled because of the high cost of revisiting sites often and the extremely high yield of these methods in terms of the number of specimens collected. Such methods would not have allowed inferences over the extent of the landscape.

In many studies of diversity, detection probability (i.e., the probability that a taxon will be scored as occuring at a site given that it was present) is implicitly assumed to be close to

one or at least uniform across taxa, but such a condition is rarely the case (Boulinier et al., 1998). Also, detection probabilities may vary among taxa, habitat type, or time (MacKenzie et al., 2006).

In the present study, where a single, small sample was taken at each site over a broad taxonomic scope, detection probabilities among families probably varied greatly. Low detection probabilies for some groups certainly influenced the results. For example, the family Acanthosomatidae (Hemiptera) is common and widespread on the landscape of the KENWR, but it was seldom collected in sweep net samples, probably because of low detection probability. Detection probability estimates could not be calculated since each site was sampled only once. This means that what was modeled was not actual family diversity, but rather the number of families collected in sweep net samples and shoud be intepreted as such.

In order to estimate detection probabilities, it is necessary to take multiple samples at each site (MacKenzie et al., 2006) or to compare multiple sampling methods. Boulinier et al. (1998); Dorazio and Royle (2005); Gelfand et al. (2003) illustrate different modelling approaches to incorporating estimates of taxon-specific detection probabilities into models of richness, but in all of these studies, multiple samples were taken at each site for estimation of detection probabilities. The issue of detection probability will be addressed more thoroughly in Chapter 2.

Chapter 2

Monitoring of Terrestrial Arthropods on the KENWR

2.1 Introduction

In this chapter I assess the use of the sampling framework of the Long-Term Ecological Monitoring Program (LTEMP) of the Kenai National Wildlife Refuge (KENWR) described in chapter 1 for monitoring of arthropod distributions on the KENWR over the long-term. Here I describe the reality of the impending re-distribution of species in response to current and projected climate change and discuss some of the ramifications for the fate of species, community assemblages, and ecosystem processes. I then introduce the proposed methods of using LTEMP to monitor for changes in species distributions using presence/absence data. I carry out analysis of proposed LTEMP sampling designs using simulated presence-absence datasets and assess the utility of these designs for monitoring distributions of arthropod species on the KENWR. I also identify characteristics of arthropods species that allowed them to be monitored well using these methods and explore the usefulness of arthropod taxa collected on LTEMP in 2004–2006 as potential monitoring metrics in the future.

Distribution shifts due to changing climate

Distribution shifts of species due to accelerated climate change are an emerging conservation concern. Mobile species are already moving up-slope and pole-ward as the climate warms (Gottfried et al., 1999; Parmesan, 1996, 2006; Parmesan et al., 1999; Parmesan and Yohe, 2003; Walther et al., 2002; Wilson et al., 2005). This is a response of organisms that have relatively stable climatic constraints to a dynamic environment. These organisms are tracking the areas that match their sets of climatic preferences, but this is not necessarily a directed or intentional response of organisms to changing climate. For example, many terrestrial

arthropods disperse more or less randomly. The net result of their random dispersal and subsequent differential success results in distribution shifts (Ashworth, 2001).

As in the past, some areas will become refugia (Gottfried et al., 1999) while others will serve as corridors (Hannah et al., 2002) as this process of re-distribution up-slope and poleward continues. An important distinction between previous climate change and the current warming is that current climate change is occurring faster than what has been typical, at least over the last few millennia (Jansen et al., 2007). While the most vagile species, such as birds, may respond rapidly, less motile species may not re-distribute themselves quickly enough to track climate. Habitat loss and fragmentation of suitable habitat will further exacerbate this problem, sometimes preventing species from moving to more suitable areas. Many species are expected to be lost as this process continues unless actions are taken to facilitate the re-distribution of species (Williams et al., 2005).

A result of this re-distribution will be the appearance of novel assemblages of species (Hannah et al., 2002). As some species shift their distributions quickly, others will respond slowly, and still others will be lost. In this process new community assemblages will form. The functional properties of these communities will be difficult to predict due to the complexity of interspecific interactions. The history of introductions of exotic species offers numerous lessons on how a change in the distribution of even a single species can have unpredictable consequences including extinctions and fundamental changes in ecosystem functioning. For example, Old World earthworms introduced into deciduous forests of northeastern North America reduce the thickness of forest litter and duff (Bohlen et al., 2004), which is detrimental to native forest plants adapted to thick duff layers (Frelich et al., 2006) and to members of the forest fauna (Migge-Kleian et al., 2006). This and other examples from our ever-growing experience with introductions of exotic species demonstrate the severity of the consequences of re-distribution of species and the appearance of new assemblages. Amelioration of the harsh climates of northern regions due to warming may increase their susceptibility to successful invasion of exotic species, further contributing to changes in community composition.

Positioned as we are at the beginning of the re-distribution of species due to accelerating climate change, the need for accurate documentation of current distributions of species and subsequent monitoring of species distributions is increasingly being recognized (Guisan and Thuiller, 2005; Magness et al., 2008).

Documenting species distributions

In its simplest form, documenting and monitoring species distribution only requires observing the presence of a species at multiple locations spread over the species' potential range. These kinds of presence-only datasets are the basis for most of the distribution information available for the world's biota. Presence-only data is suitable for establishing the limits of species' ranges.

Predictive methods based on ecological niches, such as BIOCLIM (Busby, 1991), can make use of presence-only distribution data to predict the potential distribution of species based on their observed climatic preferences. In order to do this, attributes of climate must be measured over space so that observed presences of a species can be associated with local climate characteristics. The climatic preferences of a species is inferred from the characteristics of the places where it was observed. A map of the potential range of the species can be generated by recognizing the areas on the map that fit a species' inferred climatic preferences. The limitations of these envelope methods are that they cannot express how much of a species' potential range is actually occupied by the species and that they are ill-suited for monitoring change of species distributions over time.

Including information about both where a species is present and where it is absent allows more detailed mapping of species distributions because inferences can be made about where it does and does not occur within its potential range. There are numerous methods for predicting species distributions based on presence-absence data such as regression, classification, and computer learning methods. Predictions from these kinds of species distribution models are usually expressed as probabilities of occurrence at a set of points.

Underlying these methods is an assumption of no measurement error in the response variables, i.e., a species is assumed to be truly present where it is observed and truly absent where it is not observed. As long as false negatives (i.e., failing to observe a species where it is in fact present) are unlikely, predictions of species occurrence should be essentially unbiased.

When false negatives become more likely, predictions of species distribution models will be biased, predicting probabilities of occurrence lower than the true values. Therefore the correct interpretation of these kind of predictions is that they represent a probability of observing the species (a function of both distribution and detection), which may be quite different from the species' distribution. The converse problem of false positives, usually caused by mis-identification, tends to a much smaller source of error in presence-absence data than false negatives (MacKenzie et al., 2006) (but see Royle and Link, 2006, for methods of accounting for false positives).

The issue of imperfect detection is particularly relevant in the context of monitoring. For example, if the number of sites at which a species is observed to be present increases over time, it is unclear whether (1) the species expanded its distribution or (2) the probability of detecting the species increased due to increased abundance, changes in habitat types, changes in seasonal phenology, etc. Unless detection can be assumed to be perfect (or at least constant over the time period of interest), it is necessary to account for imperfect detection in monitoring programs so that changes in the distributions can be distinguished from changes in detection probability.

An extensive literature exists on this subject of accounting for imperfect detection (see MacKenzie et al., 2006, for a review). Occupancy models explicitly account for imperfect detection using repeated survey data to obtain unbiased (or at least much less biased) estimates of presence/absence metrics (MacKenzie et al., 2003, 2006).

Occupancy models are a general class of parametric models that are used to estimate the probability of occurrence of a species at a given site. They may also be used to estimate proportion of an area or of a number of sites that is occupied by a species. Occupancy models are similar to distribution models based on logistic regression in that the parameter of interest is the probability of occurrence of a species at a site, the probability of occurrence is related by a logistic link to a function of independent variables, and maps of distributions can be generated by including spatially explicit independent variables.

Local rates of extinction and colonization (i.e., changes in distributions) are potentially some of the most useful metrics for monitoring under climate change scenarios and species introductions. These parameters can be used to monitor rates of spread of exotic species or to monitor decreasing trends in distributions of species of conservation concern while ex-

plicitly accounting for potentially dynamic detection probabilities. MacKenzie et al. (2003) presented methods of estimating these parameters using occupancy models.

To obtain data suitable for occupancy modelling, each of a set of sites is surveyed for a number of visits within a season, where a season is a time period of variable length within which it can be assumed that the target species does not change its distribution. At each site, a series of presences and absences is recorded as the site is revisited, yielding an encounter history for each site. Alternatively, encounter histories can be obtained by conducting multiple spatial surveys within a larger sampling cell. Both adequate spatial presence/absence data and encounter history data are required in order to model species distributions well and to monitor change in species distributions over time.

Monitoring species distributions on KENWR's LTEMP

In order to monitor species distributions on KENWR using the LTEMP sampling framework, I sought a sampling design, field techniques, and analytical methods most useful for monitoring of arthropod species distributions over time. I considered two proposed sampling designs, multiple sweep net samples at each site, and the use of program PRESENCE (Hines, 2007) for analysis.

LTEMP was designed to inventory, model distributions of, and monitor the flora and fauna of KENWR over the long-term. A systematic grid of 327 points was distributed over the KENWR at 4.8km intervals, of which 255 points did not fall on water or ice. Over the field seasons of 2004–2006, all the 255 points were sampled for vascular and non-vascular plants, breeding birds, and arthropods. The background, sampling design, and field methods of LTEMP are described in more detail in Chapter 1.

The large sample size and nearly complete spatial coverage of LTEMP provided ample presence/absence data for documenting species distributions on the KENWR. For example, Magness et al. (2008) successfully modelled distributions of multiple bird species on the KENWR. However, since each site was sampled only once in 2004–2006, detection probabilities could not be estimated, so it was not possible to produce unbiased estimates of species distributions. For this reason the models of arthropod family diversity in Chapter 1 are certainly underestimates because many of the arthropods included in that analysis are commonly missed in sweep net samples (pers. obs.).

In the future, LTEMP methods will account for imperfect detection. The most common method of obtaining detection histories required for occupancy modelling is to visit each site on multiple occasions within a season. However, making multiple visits is not compatible with LTEMP methods. Since all sites are sampled by helicopter, the cost of revisiting the sites within a season would quickly become cost-prohibitive. In addition, since LTEMP is designed as a long-term monitoring regime, it is desirable to minimize damage to the sites. Raising the number of visits would cause additional trampling and increase the likelihood of introducing exotic species. For these reasons, KENWR's managers plan to employ methods that allow estimation of detection probabilities within single visits. For arthropods, I propose taking two sweep net samples on each site. The 5.4m radius circular plots described in Chapter 1 would be split in half along a north-south line, and a sweep net sample would be taken over each semicircle. These two samples would constitute two spatial samples within the larger plot that could be used to estimate probabilities of detection for each species sampled.

After the complete census of all of the LTEMP sites over 2004–2006, the original memorandum of understanding (MOU) between the US Forest Service's Forest Inventory and Analysis (FIA) program KENWR, which outlined LTEMP, called for sampling 20% of the 255 LTEMP sites every two years. This is a rotating panel design (interpenetrating panel design) where each of the five panels is well-distributed across and representative of the KENWR (Table 2.1). The advantages of the rotating panel design are (1) concurrent sampling by FIA and LTEMP field crews provides optimal temporal proximity of the components of the dataset; (2) each panel is representative of the KENWR as a whole; (3) estimates of variables of interest are obtained on a frequent basis; and (4) over time, the sample size is large with excellent spatial coverage, enabling modelling and monitoring of species distributions. The main disadvantage of this design is potentially poor precision of yearly estimates due to the relatively small number of sites sampled each season. However, methods exist for improving precision of estimators from rotating panel data by incorporating information from previous seasons (Van Deusen, 1999, 2002).

A proposed alternative sampling design requiring the same overall amount of sampling effort would be to perform a periodic census of all sites once every ten years (Table 2.1). In the sample years, this design would yield more accurate estimates of parameters of interest

than the rotating panel design because its sample size would be larger, but the cost would be no information obtained for the nine intervening years.

Table 2.1: Tabular representation of proposed LTEMP sampling designs over 20 years. Each row represents a panel of 51 sites. X's (X) represent plots surveyed; dashes (-) indicate times when sites were not surveyed.

Year	0	2	4	6	8	10	12	14	16	18
Design 1	X	-	-	-	-	X	-	-	-	-
Rotating Panel	-	X	-	-	-	-	X	-	-	-
	-	-	X	-	-	-	-	X	-	-
	-	-	-	X	-	-	-	-	X	-
	-	-	-	-	X	-	-	-	-	X
Design 2	X	-	-	-	-	X	-	-	-	-
Periodic Census	X	-	-	-	-	X	-	-	-	-
	X	-	-	-	-	X	-	-	-	-
	X	-	-	-	-	X	-	-	-	-
	X	-	-	-	-	X	-	-	-	-

At present, a few software packages are available for fitting occupancy models: PRESENCE , MARK (White and Burnham, 1999), and WinBUGS (Thomas, 1994; Spiegelhalter et al., 2003). Of these options, KENWR's managers have proposed using PRESENCE for analysis of LTEMP occupancy because there is a good precedence with use of this software in occupancy literature; because it is quite flexible, allowing for diverse designs and model types; and because it is the most intuitive of the available software options. This latter concern is particularly important for a long-term program such as LTEMP, where there may be turnover of personnel during the course of sampling. In this situation there is strong motivation for making analyses as simple and consistent as possible, which is the main reason KENWR's managers selected PRESENCE for estimation of occupancy metrics on LTEMP.

However, analyses will not necessarily be limited to program PRESENCE. The data obtained from the proposed designs would be suitable for monitoring and modelling using multiple methods. For example, a hierarchical model could be written that (1) models both occupancy and detection as functions of covariates, (2) includes geospatially autocorrelated

error terms, and (2) models the Markovian processes of extinction and colonization (see Royle and Kéry, 2007; Royle et al., 2007, for examples).

Objectives

Objective 1: Evaluate proposed LTEMP sampling regimes for monitoring arthropod species using occupancy methods. My first goal was to compare the performance of the two proposed LTEMP sampling designs for obtaining unbiased species distributions (site-specific estimates of occupancy, the probability of occurrence of a species at a site) within a single season. My second goal was to compare the ability of these sampling designs to estimate local rates of colonization over the short-term (sampling of two consecutive panels over four years) and the long-term (two complete surveys of all the LTEMP panels over 20 years).

Objective 2: Evaluate arthropod taxa collected in LTEMP sweep net samples as potential monitoring candidates. Given the performance of the LTEMP sampling designs for estimating occupancy metrics over time and the observed frequencies of arthropod taxa in LTEMP samples, I sought to discern which arthropod taxa could be monitored well using LTEMP methods.

2.2 Methods

Evaluating LTEMP monitoring designs

My general approach for evaluating LTEMP sampling regimes was to assess the performance of the proposed designs through Monte-Carlo simulation. These simulations were designed to answer the question of whether or not occupancy metrics could be estimated well; they were not designed to actually model species distributions. In the same way that a mean can be considered a special case of linear regression where there is only an intercept and an error term, the simple occupancy models I used estimated only occupancy metrics (occupancy, detection probability, and, in multi-season cases, rates of colonization and extinction) without considering additional variables (e.g., possible covariates) that would normally be included to produce species distribution models. These simulations were designed to answer the questions of interest without unnecessary complexity.

For each scenario considered, I (1) generated large numbers of simulated datasets that conformed to proposed LTEMP monitoring designs and had known parameter values (2) fitted occupancy models using program PRESENCE, and (3) compared the estimates obtained by program PRESENCE to the known parameter values. I wrote R (R Development Core Team, 2006) scripts to automate the process of generating datasets, writing input files for program PRESENCE, executing PRESENCE over the command line, extracting results from PRESENCE's output files, and summarizing the results of large numbers of simulations.

Note that the way in which simulated data were generated and the way in which models were subsequently fitted were appropriate for estimating the population parameters as opposed to the finite sample parameters, i.e., the plots at which observations were made were a sample from a larger population, and estimates yielded would be estimates of population parameters. This was in contrast to estimating occupancy of a finite population of plots.

Generating simulated data

I made datasets suitable for occupancy modeling using R. For a given set of scalar values of occupancy (Ψ), detection probability (p), number of sites (n), and number of surveys (K), I first generated a list of occupancy states at n sites by specifying that the occupancy states at all sites were independently and identically distributed realizations of a Bernoulli process with a rate of Ψ. I used the `rbern()` function in the `Rlab` add-on library (Boos et al., 2006) to do this. Similarly, I specified that detections on each of the K revisits at the n sites were independently and identically distributed realizations of a Bernoulli process with a rate of p. Multiplying the detection states by the occupancy states yielded detection histories with the specified parameter values of Ψ, p, n, and K. An example of an R function for generating a simulated dataset is provided below.

```
### Function for generating simulated detection histories.
   generate.data <- function(psi, p, n, K)
    {
    ### Load the Rlab package.
    require(Rlab)
    ### Make an array of the right size to hold the data.
    data <- array(data=NA, dim=c(n,K))
    ### Generate occupancy states at each site.
    z <- rbern(n, psi)
    for (i.K in 1:K)
     {
     ### Generate detection states at each site.
     b <- rbern(n, p)
     ### Multiply the occupancy states by the detection states.
     data[,i.K] <- z*b
     }
    ### Return the data.
    data
    }
```

Executing this function yields data of the form where rows represent sites and columns represent surveys. An example of the execution of this function is provided below.

```
> generate.data(psi=0.5, p=0.8, n=5, K=4)
      [,1] [,2] [,3] [,4]
[1,]    1    1    0    1
[2,]    0    0    0    0
[3,]    0    1    1    1
[4,]    1    1    1    1
[5,]    0    0    0    0
```

This is essentially the function used except that I modified it to generate large numbers of datasets for multiple values of Ψ and p.

Multi-season data were generated by simulating the Markovian processes of colonization and extinction over a specified number of seasons. For each site that was occupied at season t, there was a probability ε that it would become unoccupied by season $t+1$; sites unoccupied in season t had a probability γ of being colonized by season $t+1$. The R code I used for generating multi-season occupancy states is included below. In this code, `occupancy` is an array of occupancy states with dimensions `n` (number of sites) and `seasons` (number of seasons).

```
occupancy <- array(data=NA, dim=c(n,seasons))
   occupancy[,1] <- rbern(n, psi)
   for (i.season in 2:seasons)
    {
    for (i.n in 1:n)
     {
     ### Colonization
     if (occupancy[i.n,i.season-1] == 0)
      {
      occupancy[i.n,i.season] <- rbern(1,gamma)
      }
     ### Extinction
     if (occupancy[i.n,i.season-1] == 1)
      {
      occupancy[i.n,i.season] <- 1 - rbern(1,epsilon)
      }
     }
    }
```

Obtaining naïve estimates of occupancy

Naïve estimates of occupancy were calculated by counting the sites where a species was observed on any survey and dividing that value by the total number of sites.

Fitting occupancy models using program PRESENCE

I fitted occupancy models through R code that generated input files from the simulated datasets and invoked program PRESENCE over the command line. For each simulated dataset, data and model structure text files of the format required by program PRESENCE were created. PRESENCE was executed over the command line with a command like

```
> presence.exe i=input.pao l=output.pa2 name="model name" model=11c lmt=100 %
      N=254 T=2 NPar=2
```

where `i` is the input file, `l` is the output file, `name` is a name for the model, `model` describes model structure, `lmt` tells PRESENCE whether not to use a single season (`lmt=100`) or multi-season (`lmt=200`) model, `N` is the number of sites, `T` is the number of surveys ($\texttt{T} = K$), and `NPar` is the number of parameters to be estimated.

Results were extracted from the output files produced by these commands. The R scripts used for generating simulated datasets, fitting occupancy models, and retrieving the results are given in Appendix D.

Single-season scenarios

In order to assess the performance of the rotating panel and periodic census designs for obtaining estimates of single-season occupancy states, I ran simulations under both proposed sampling regimes. The only difference between these designs for a single season was the sample size: $n = 51$ for the rotating panel design and $n = 254$ for the periodic census design. I considered nine values of occupancy (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9) and the same nine values of detection probability. For each combination of parameters, I generated 100 datasets (a total of $9^2 \times 100 = 8{,}100$ datasets). Using the same simulated datasets, I obtained both naïve estimates of occupancy and estimates from program PRESENCE. I used an occupancy model that assumes constant detection probability over time and across all sites.

Short-term, multi-season scenarios

In multi-season scenarios, the number of variables (4) becomes too large to explore much of the parameter space because the number of permutations quickly becomes intractable (using similar methods as the single-season analysis, this would be (9 values)^4 parameters $\times$ 100 simulations $= 656,100$ simulations). In order to assess how the the rotating panel design would perform over the short term (2 seasons or 4 years) for estimating local rates of colonization and extinction, I chose to examine how a realistic pair of values of these change parameters would be estimated over a wide range of values of Ψ and p. I set the local rate of extinction to 0.1, the local rate of colonization to 0.05, and the number of seasons to two. I considered the same set of values of Ψ and p used in the single-season simulations above. Since sampling would take place every two years for the rotating panel design, the definition of a season for this design was a two-year span.

For these and all other multi-season simulations, I chose a simple multi-season occupancy model which considered p, ε, and γ to be constant over time (equation 2.1). In this model, only the initial value of Ψ, Ψ_1, is directly estimated; subsequent values can be estimated by plugging values of Ψ_1, ε, and γ into equation 2.1.

$$\Psi_{t+1} = \Psi_t + (1 - \Psi_t\gamma) - \Psi_t\varepsilon \tag{2.1}$$

The periodic census design would yield no information over the short-term, so no short-term analysis of this design was run.

Long-term, multi-season scenarios

Over the long-term, I was most interested in assessing how well the change parameters ε and γ would be estimated. I wanted to know how long it would take for each design to yield good estimates and to find the accuracy (or inaccuracy) of these estimates. I simulated implementation of both designs over 20 years, two complete sweeps of the sample frame.

For the rotating panel design, I intentionally picked parameter values that were realistic and easily estimated. I set initial Ψ to 0.7, p to 0.8, ε to 0.1, and γ to 0.05. I avoided using values at the extremes of parameter space of Ψ and p, which likely would have led to a high failure rate. I gave p a high value since single-season analyses demonstrated that these occupancy models performed worst when detection probability was low. I picked realistic values of ε and γ for a population in decline over 20 years.

Since the definition of a season differs between the two designs, they might yield quite different values of ε and γ from the same populations over the same amount of time. In the rotating panel design, ε is defined as the probability of local extinction over two years; in the periodic census design, ε is the probability of local extinction over ten years. The latter would always be expected to be larger than the former if ε for each year is greater than zero. In order to compare the designs realistically, I calculated the values of ε and γ equivalent to applying the values in the above analyses over the longer, 10-year season of the periodic census design. I set $\Psi = 0.7$, $p = 0.8$, $\varepsilon = 0.3709$, and $\gamma = 0.1854$.

Dealing with convergence failures

Throughout the simulation analyses, program PRESENCE often yielded fitted values of zero or one for either detection probability or occupancy parameters. When this occurred, it appeared as if one parameter was fixed at either zero or one while fitting was performed on the remaining variable, leading to erroneous results. I considered these instances failures. Program PRESENCE also considered these instances failures, issuing the warning "numerical convergence was not reached" in its output.

The failure rate of a set of simulations was calculated by dividing the number of failed simulations by the total number of simulations, expressed as a percentage. All other summaries excluded results from simulations where PRESENCE failed to converge, summarizing only simulations where PRESENCE converged successfully.

Evaluating Arthropod Taxa as Monitoring Candidates.

In order to determine which of the arthropod taxa collected in 2004–2006 could be monitored well over the long term, I compared inferred occupancy and detection probabilities of these taxa to the thresholds of occupancy and detection probability necessary for a taxon to be monitored well on LTEMP. These thresholds were obtained from the above simulation analyses.

Since occupancy metrics could not be directly estimated from 2004–2006 LTEMP data, I was limited to inferring occupancy characteristics of arthropod taxa using their observed frequencies. Because observed frequency is essentially the product of occupancy and detection probability, it could be inferred that taxa with high observed frequencies also had high occupancy and high detection probability. However, low observed frequencies were ambiguous because this could have been caused by low occupancy, low detection probability, or both. For this reason, taxa with low occupancy and high detection probability (which could be monitored with good accuracy) could not be distinguished from taxa with low detection probability (which could not be monitored with acceptable accuracy).

2.3 Results

Evaluation of LTEMP monitoring designs

Single-season estimates

Program PRESENCE had a high failure rate, especially when detection probability (p) was low (Figures 2.1 and 2.2, Tables C.1 and C.2). The rotating panel design had a much higher failure rate, with an average of 24 failures out of every 100 simulations, than the periodic census design, where an average of eight of every 100 simulations resulted in failures. The failure rate was always highest (up to 98 failures out of 100 simulations for the rotating panel design) when both occupancy (Ψ) and detection probability (p) were lowest. Program

PRESENCE was most reliable when p was between 0.6 and 0.8 and Ψ was between 0.3 and 0.7.

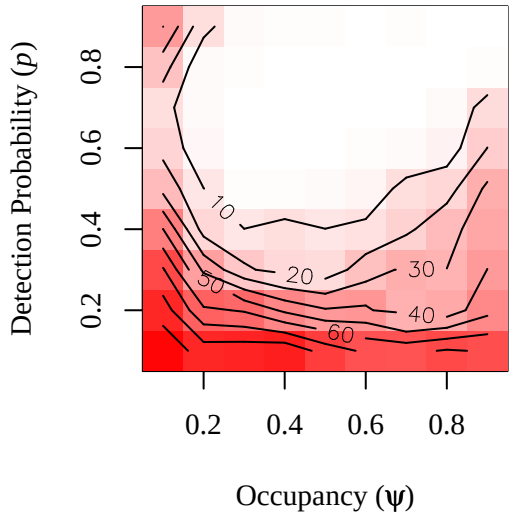

Figure 2.1: Failure rate of Program PRESENCE over a range of values of Ψ and p for the rotating panel design. Values are the percentage of simulations in which PRESENCE failed. Failures were defined as fitting of either Ψ or p with a value of 1 or 0. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 100%.

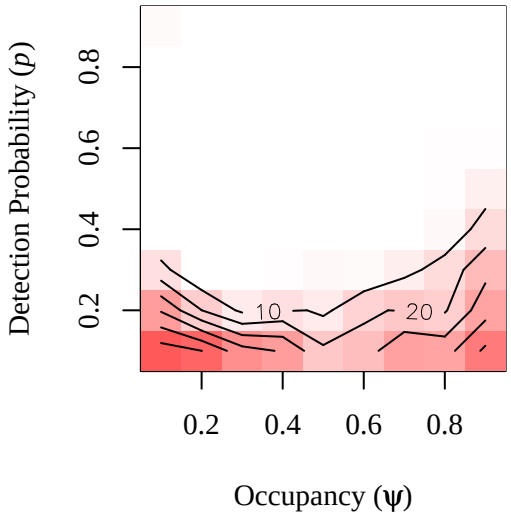

Figure 2.2: Failure rate of Program PRESENCE over a range of values of Ψ and p for the periodic census design. Values are the percentage of simulations in which PRESENCE failed. Failures were defined as fitting of either Ψ or p with a value of 1 or 0. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 100%.

Naïve estimates of Ψ were nearly always biased low, with bias increasing as p decreased (Figures 2.3 and 2.4, Tables C.3 and C.4). Only when detection probability was highest (0.9) were naïve estimates nearly unbiased.

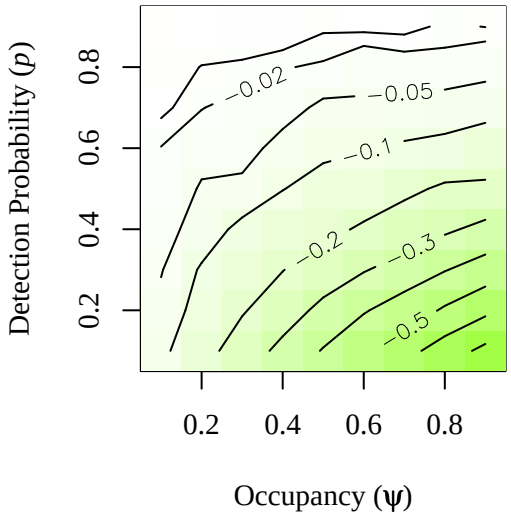

Figure 2.3: Bias of naïve estimates of Ψ for the rotating panel design. The colors of the pixels are graduated so that white represents a value of 0, saturated red represents a value of 1, and saturated green represents a value of -1.

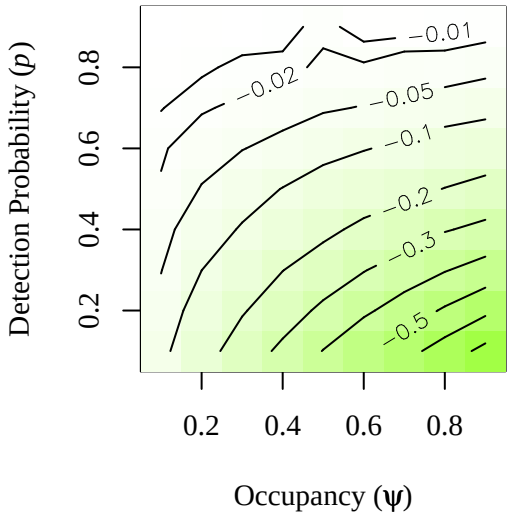

Figure 2.4: Bias of naïve estimates of Ψ for the periodic census design. The colors of the pixels are graduated so that white represents a value of 0, saturated red represents a value of 1, and saturated green represents a value of -1.

When Program PRESENCE did not fail, it produced generally unbiased estimates of occupancy as long as detection probability was greater than 0.3 (Figures 2.5 and 2.6, Tables C.5 and C.6). When p was less than 0.3, estimates of Ψ tended to be biased low. The magnitude of bias was greatest when both Ψ and p were extremely low (0.1). At this extreme of the parameter space, estimates of Ψ were highly variable when PRESENCE did not fail. Program PRESENCE nearly always yielded less biased estimates of occupancy than the naïve estimates.

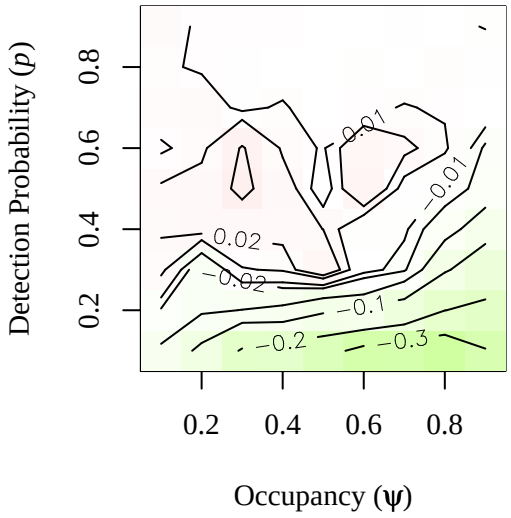

Figure 2.5: Bias of PRESENCE estimates of Ψ for the rotating panel design. The colors of the pixels are graduated so that white represents a value of 0, saturated red represents a value of 1, and saturated green represents a value of -1.

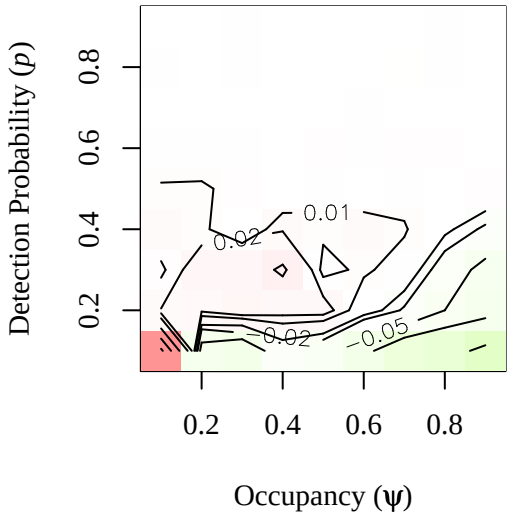

Figure 2.6: Bias of PRESENCE estimates of Ψ for the periodic census design. The colors of the pixels are graduated so that white represents a value of 0, saturated red represents a value of 1, and saturated green represents a value of -1.

Standard deviations of all estimates of occupancy were higher for the rotating panel design (Figure 2.7, Table C.9) than for the periodic census design (Figure 2.8, Table C.10) For the rotating panel design, standard deviation was nearly always greater than 0.05; for the periodic census design, standard deviation was generally less than 0.05 when detection probability was greater than 0.6. In both cases, the value of occupancy had relatively little influence on precision, which was determined mostly by detection probability. Estimates of occupancy obtained from program PRESENCE were generally much less precise than naïve estimates (Figures 2.9 and 2.10).

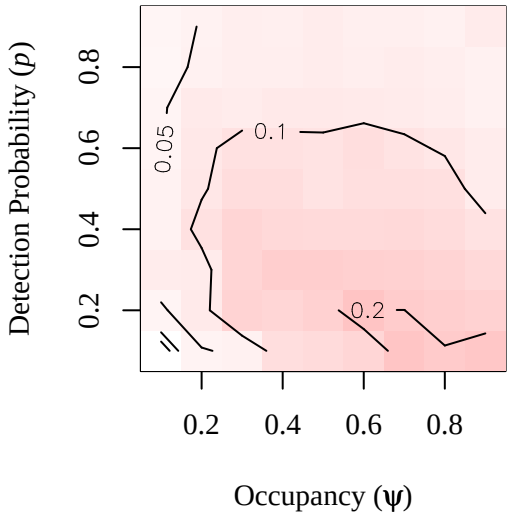

Figure 2.7: Standard deviation of PRESENCE estimates of Ψ for the rotating panel design. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 1.

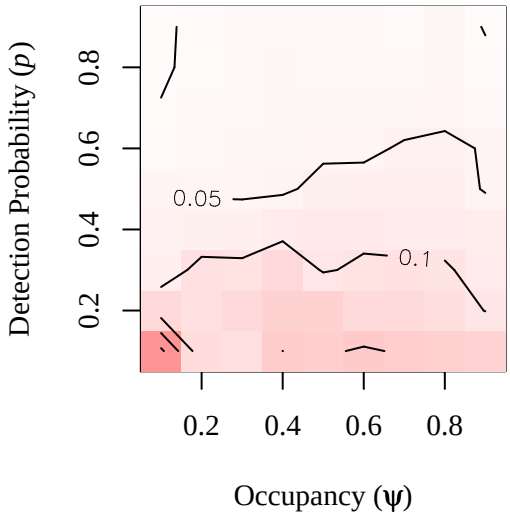

Figure 2.8: Standard deviation of PRESENCE estimates of Ψ for the periodic census design. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 1.

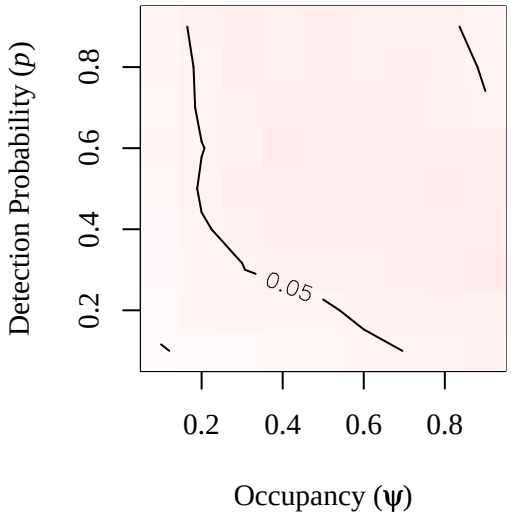

Figure 2.9: Standard deviation of naïve estimates of Ψ for the rotating panel design. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 1.

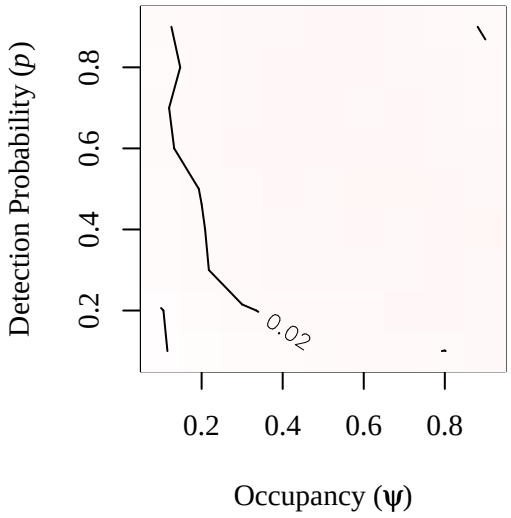

Figure 2.10: Standard deviation of naïve estimates of Ψ for the periodic census design. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 1.

Multi-season estimates: short term performance

For the rotating panel design, local rates of extinction and colonization could not be estimated well over the short term. The failure rate of program PRESENCE was greater than 10% whenever detection probability was 0.3 or lower, but its failure rate was generally negligible when detection probability was higher (Figure 2.11, Table C.11).

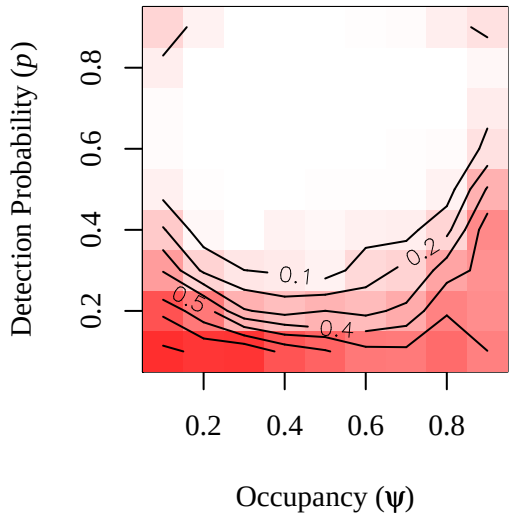

Figure 2.11: Failure rate of program PRESENCE over 4 years for the rotating panel design. Failures were defined as fitting of either Ψ or p with a value of 1 or 0. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 100%.

Estimates of the local rates of colonization and extinction were initially biased high (Figures 2.12 and 2.13) and standard deviation of the estimates was essentially proportional to the bias (Figures 2.14 and 2.15; Tables C.12 and C.14). Bias of ε ranged from 0.091 to 0.669; bias of γ ranged from 0.006 to 0.912.

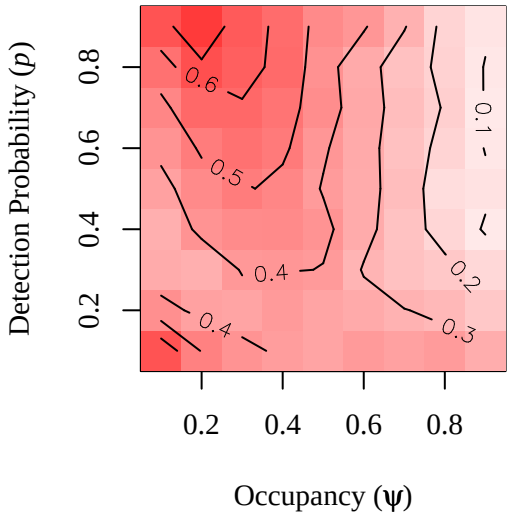

Figure 2.12: Mean bias of estimates of the rate of local extinction for the rotating panel design over four years. The colors of the pixels are graduated so that white represents a value of 0, saturated red represents a value of 1, and saturated green represents a value of -1.

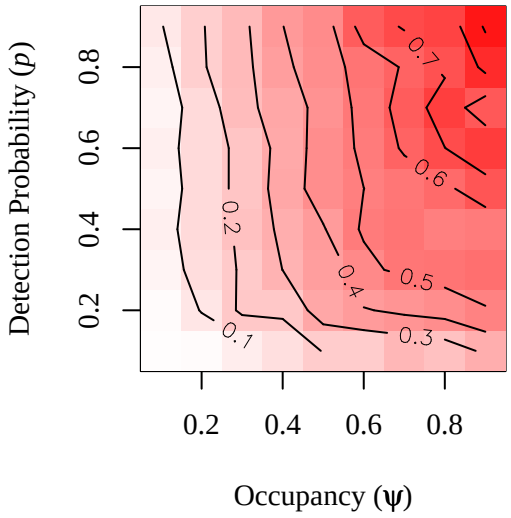

Figure 2.13: Mean bias of estimates of the local rates of colonization for the rotating panel design over four years. The colors of the pixels are graduated so that white represents a value of 0, saturated red represents a value of 1, and saturated green represents a value of -1.

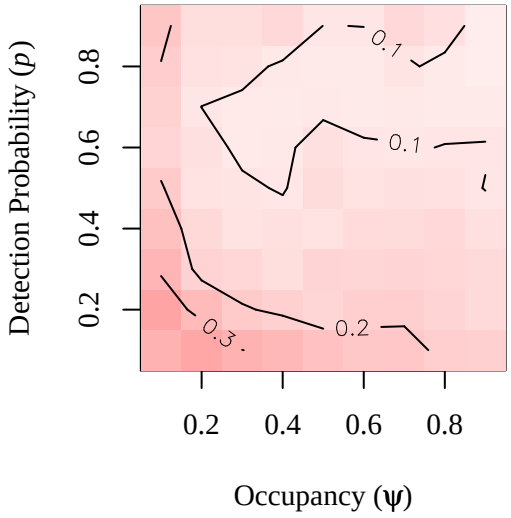

Figure 2.14: Standard deviation of the local rate of extinction for the rotating panel design over four years. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 1.

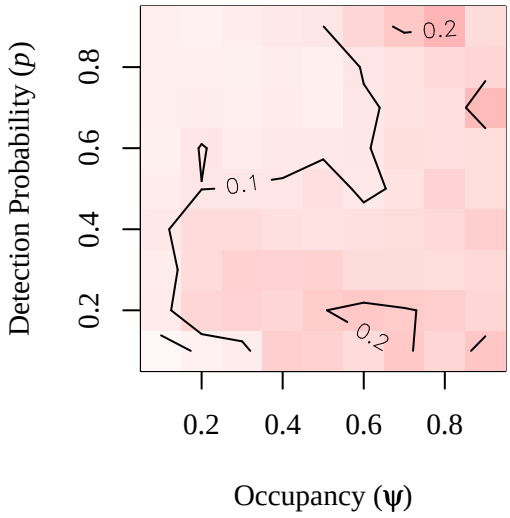

Figure 2.15: Standard deviation of estimates of the local rates of colonization for the rotating panel design over four years. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 1.

The periodic census design would provide no information over consecutive seasons.

Multi-season estimates: long term performance

The rotating panel design did not yield reasonable estimates of local rates of colonization and extinction until some sites were visited twice, ten years into the sampling program. After this point, estimates of these change parameters were generally quite good. Estimates of occupancy were consistently unbiased (mean bias = 0.004), but not quite as precise as the desired maximum standard deviation of about 0.05 (SD = 0.062).

Over time, the failure rate of program PRESENCE first rose dramatically from 0 at the second year to 67% at the fourth year, then remained high until the tenth year, when some sites were sampled a second time (Figure 2.16). After this point, the failure rate became negligible.

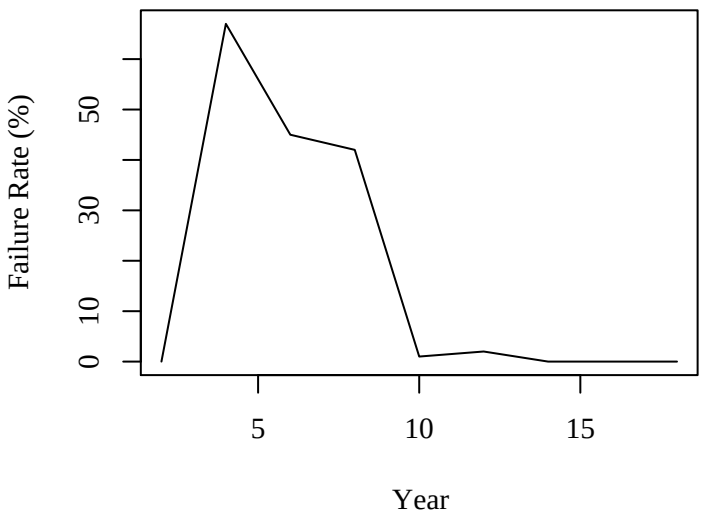

Figure 2.16: Failure rate of program PRESENCE over 20 years of the rotating panel design. Values are the percentage of simulations in which PRESENCE failed. Failures were defined as fitting of Ψ, p, ε, or γ with a value of 1 or 0.

Consistent with the short-term analysis above, estimates of the change parameters ε and γ were initially biased high (Figures 2.17 and 2.18). Mean bias at the second year was 0.25 for ε and 0.61 for γ. This bias diminished slowly until the tenth year, when bias dropped quickly as some sites were sampled a second time. By the twelfth year and subsquently, estimates of these change parameters were essentially unbiased. After year twelve, precision was still not as good as is desirable, with standard deviations of about 0.1.

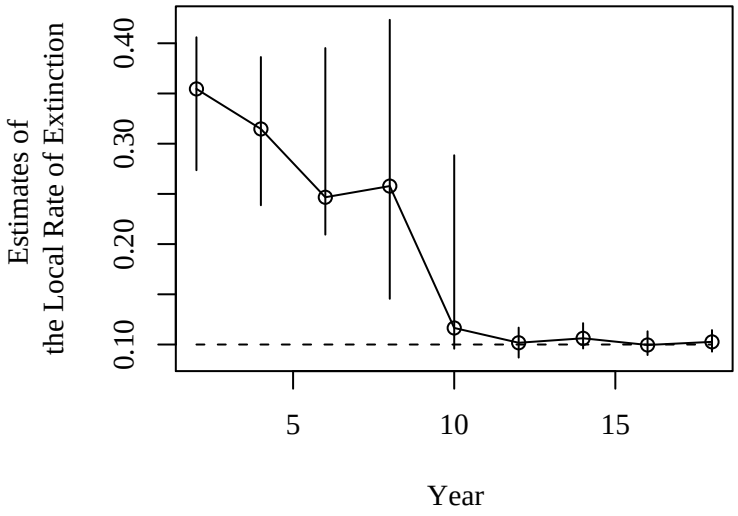

Figure 2.17: Estimates of the local rate of extinction over 20 years for the rotating panel design. Dashed line: true value (0.1). Circles and solid line: median of estimates with bars spanning from 25% to 75% quantiles.

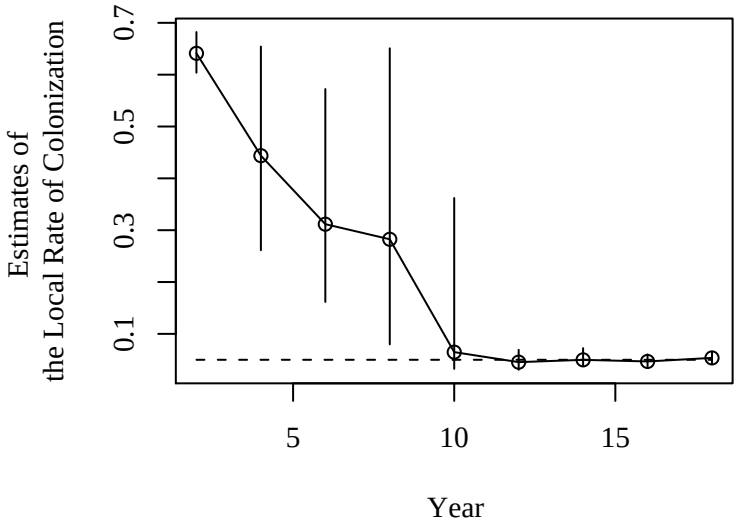

Figure 2.18: Estimates of the local rate of colonization over 20 years for the rotating panel design. Dashed line: true value (0.05). Circles and solid line: median of estimates with bars spanning from 25% to 75% quantiles.

The periodic census design yielded accurate estimates of local rates of colonization and extinction (Figures 2.19 and 2.20). Estimates of both parameters were were unbiased (mean bias was -0.0004 for ε and -0.0006 for γ) and acceptably precise (standard deviation was 0.041 for ε and 0.051 for γ). There were no convergence failures for this scenario.

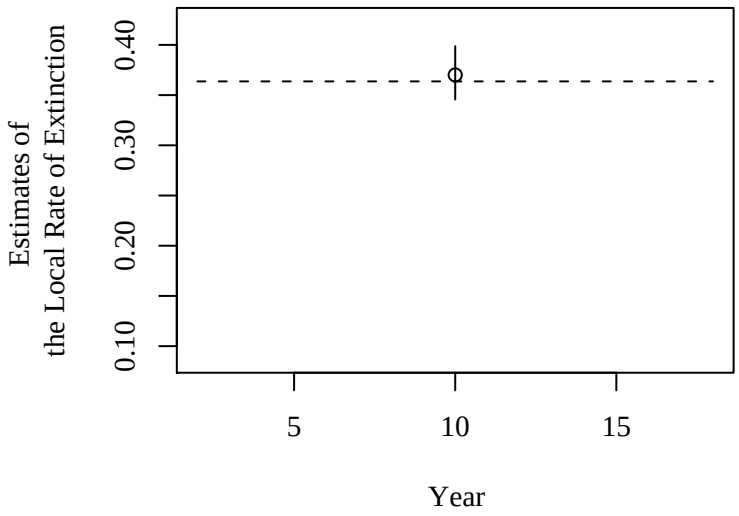

Figure 2.19: Estimates of the local rate of colonization over 20 years for the periodic census design. Dashed line: true value (0.371). Circle: median of estimates with bars spanning from 25% to 75% quantiles.

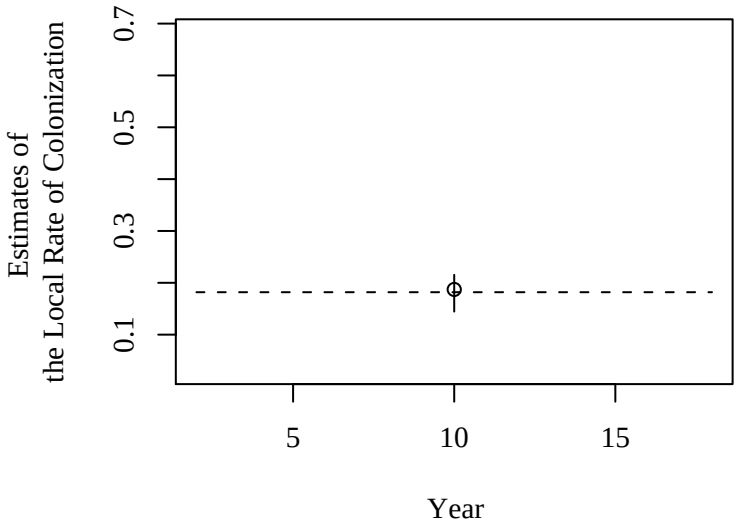

Figure 2.20: Estimates of the local rate of colonization over 20 years for the periodic census design. Dashed line: true value (0.185). Circle: median of estimates with bars spanning from 25% to 75% quantiles.

Evaluation of arthropod taxa as monitoring candidates.

A minimum threshold of detection probability of about 0.5 was necessary for occupancy to estimated with acceptable precision (standard error of about 0.05) for the single-season estimates from the periodic census design (Figure 2.8), which is comparable to estimates of occupancy yielded by either design over this term.

Of the families encountered, Aphididae, Braconidae, Culicidae, Ichneumonidae, Muscidae, Psyllidae, Simuliidae, and Sminthuridae were collected at $\geq 22\%$ of sites. It is probable that these groups had detection probabilities near 0.5, so they may be monitored well on LTEMP. Culicidae, Aphididae, and Sminthuridae were the most frequently collected families, found on 76%, 60%, and 51%, respectively, of plots. It is almost certain that these families could be monitored with good precision and accuracy on LTEMP since it is extremely unlikely that their detection probabilities are less than their observed frequencies.

Of the species identified on LTEMP at this point, only *Sminthurus* A (Christiansen and Bellinger, 1998) (Sminthuridae), can be recognized as having high enough observed frequency to infer that its detection probability must be high enough that it can be successfully studied.

A Delphacid planthopper common on the KENWR, *Javesella pellucida* (Fabricius, 1794), also had high detection probability (about 0.8) in sweep net samples similar to LTEMP methods conducted in a small area on KENWR (unpublished data), so monitoring this species should be possible using LTEMP methods. Species identifications have not yet been made for most of the specimens represented in 2004–2006 LTEMP samples. The list of species that can be monitored well using LTEMP methods will grow as more species are identified.

2.4 Discussion

Evaluation of LTEMP monitoring designs

Based on my analyses, I recommend the rotating panel design over the periodic census design for monitoring of species distributions on the Kenai National Wildlife Refuge. While the periodic census offers greater accuracy, the rotating panel has the notable advantage of yielding more frequent and up-to-date information on colonization and extinction. These data will be critical for management decisions under climate change scenarios.

The main advantage of the rotating panel design is that it would enable frequent updating of species distribution models and it would provide up-to-date estimates of colonization and extinction. The only disadvantage of the rotating panel design is that it is not as precise as the periodic census design, but it would be possible to improve precision of this design by increasing the number of sweep net samples taken at each site. In occupancy models, increasing the number of surveys is generally more effective for increasing precision than increasing the number of sites by the same factor, especially when detection probability is low (MacKenzie and Royle, 2005; MacKenzie et al., 2006). Increasing the number of sweep net samples per site would certainly be more effective than increasing the number of sites for sampling arthropods on LTEMP, where the number of surveys is at its minimum value for estimating detection probabilities ($K = 2$), sample size is already large ($n = 254$), and detection probabilities for most arthropod species are low. A statistical method that may improve precision of the rotating panel design would be the application of mixed estimation methods for estimation of parameters from rotating designs (Van Deusen, 1996, 1999, 2002) to occupancy modelling, but this option has not been explored. For arthropods, accuracy of

occupancy metrics may also be improved by incorporating information on relative abundance as was done by Royle and Nichols (2003) since the sweep net samples yield data on relative abundance. Because of these options for improving precision, the advantage of providing up-to-date information far outweighs the rotating panel's main disadvantage of inferior precision compared to the periodic census design.

Neither monitoring design provided good estimates of short-term processes. Rates of local extinction and colonization could not be accurately estimated because individual sites were not resampled in consecutive years. Using the rotating panel design as proposed above, a more appropriate parameterization for the first eight years of monitoring would be to estimate a simple rate of change in occupancy rather than rates of colonization and extinction. Alternatively, the rotating panel sampling design could be augmented with a set of additional sites that would be sampled in consecutive seasons with the explicit purpose of providing up-to-date estimates of rates of colonization and extinction.

A final argument recommending the rotating panel design over the periodic census design is that data obtained from the rotating panel design can be binned by decade to be analyzed in the same way as data from the periodic census design, achieving similar levels of precision. In contrast, there is no way to analyze data from the periodic census design so that it acquires the rotating panel's advantage of providing frequently-updated information.

Software and failure rates

Program PRESENCE, as the proposed software package for monitoring occupancy on LTEMP, would be suitable for monitoring occupancy metrics for species with detection probabilities ≥ 0.5. For species with lower detection probabilities or extremely low or high occupancy, PRESENCE would generally not be satisfactory due to high failure rates. Fortunately, PRESENCE issues warnings whenever these convergence failures occur, alerting the user that results may not be reasonable. PRESENCE appears to be generally accurate whenever convergence failures do not occur. (When I compared running some simple occupancy models in MARK and PRESENCE, they always agreed to within rounding error, and they had identical failure rates.)

A deficiency of PRESENCE for monitoring species distributions is that it is not predictive. It is useful for estimating occupancy parameters and for hypothesis testing, but it has no

built-in methods for making predictions over the landscape. In order to make distribution maps from PRESENCE models, that portion of an occupancy model describing occupancy as a function of covariates would have to be applied over the landscape using a geographic information system (GIS) or programming language such as R.

For situations in which PRESENCE has a high failure rate, I recommend the use of Bayesian occupancy models in WinBUGS, such as the code presented by MacKenzie et al. (2006) and Royle and Kéry (2007). In my experience using these kinds of models, they yield sensible estimates of occupancy metrics even at the edges of parameter space or when sample size is small. In addition, a map of predictions with credibility intervals can be generated. The flexibility of WinBUGS also allows for explicitly spatial modelling, for example with geostatistically correlated error terms (e.g., Royle and Kéry, 2007).

Because of its ease of use, I recommend PRESENCE for monitoring of occupancy parameters and modelling distributions on LTEMP whenever it works well; wherever PRESENCE tends to fail to converge, I recommend running occupancy models in WinBUGS because it is less susceptible to this problem.

Evaluation of arthropod taxa as monitoring metrics

Any readily identifiable taxon that is not especially rare and is likely to be collected in sweep net samples should be monitored well by LTEMP. Since detection probabilities could not be estimated well from the 2004-2006 LTEMP dataset, choosing potential monitoring candidates is problematic for all but the most frequently collected taxa. Taxa that may have had high detection probabilities but low values of occupancy could not be distinguished from taxa with low detection probabilities. The families which were frequently encountered necessarily had detection probabilities at least as high as the observed frequencies, which means that occupancy metrics for these groups could be estimated well given the proposed sampling regimes.

The potential usefulness of the families Aphididae, Braconidae, Culicidae, Ichneumonidae, Muscidae, Psyllidae, Simuliidae, and Sminthuridae, all of which had fairly high estimated detection probabilities in sweep net samples, is discussed below.

Aphididae Aphididae feed on juices of diverse plant hosts, including most deciduous trees and shrubs, conifers, and herbs. Aphid populations may fluctuate wildly, with short generation times and high fecundity allowing explosive population growth under favorable conditions (Dixon, 1977). Dense aphid infestations can adversely affect growth and reproduction of host plants (Dixon, 1985). They are also vectors of plant viruses (Borror et al., 1989). Aphids are important prey species for many arthropod predators and parasitoids, including members Coccinellidae, Hemerobiidae, Chrysopidae, Coniopterygidae, Braconidae, Ichneumonidae, Syrphidae, and Chamaemiidae.

Although the family Aphididae were easily sorted from among the other arthropods, they were often damaged, frequently immature, and difficult to identify to species. Aphididae are also generally monophagous, each species feeding on a narrow range of host plants (Dixon, 1985), which would make information about Aphididae somewhat redundant with vegetation data. For these reasons, potential monitoring of Aphididae would be most reasonable at the coarse family resolution. Abundance of Aphididae may be an informative monitoring metric because of their damage to plant hosts and because of their importance as prey for many arthropod predators and parasites.

Braconidae Braconidae are parasitoids on diverse insects. Many of the Braconids represented in LTEMP sweep net samples were parasitoids of aphids.

Most Braconidae collected were identified to genera, and many were identified to species by Dr. Michael Sharkey (University of Kentucky) and Dr. Petr Stary (Czech Academy of Sciences, Czech Republic). Many of the identifications only to genera were due to taxonomic issues that have yet to be resolved. No species was collected frequently enough to be monitored well. Although monitoring of particular species with known parasitoid-host relationships may be informative (particularly because many are parasites of Aphididae), monitoring of Braconidae at the family resolution does not appear to be justified.

Cicadellidae Cicadellidae feed on juices of various vascular plants, mostly grasses, shrubs, and herbs (Beirne, 1956). Some of the most common species present on the KENWR (e.g., *Balclutha punctata*) are generalists as adults, feeding on a variety of native and introduced grasses. Cicadellidae are also vectors of plant diseases (Borror et al., 1989).

Cicadellidae were easily sorted from among other families, and they tended not to be damaged; but consistent species determinations would not be possible because many of the LTEMP specimens of Cicadellidae were immature. Identifications of Cicadellidae to species generally requires examination of adult males. Because they can often not be identified, Cicadellidae may be best monitored at the family resolution. Since they occupy a similar ecological niche as the Aphididae, information about the relative abundances of these groups over time could be interpreted similarly.

Delphacidae Delphacids are sap-feeders, most of them feeding on graminoids. Most are monophagous (Wilson, 1997), but some are generalists feeding on a variety of graminoid species. Although many immature Delphacids were collected, adult males collected in LTEMP sweep net samples were easily identified using available keys (e.g., Wilson, 1988) and through the help of Dr. K. G. A. Hamilton (Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada). By far the most abundant species of Delphacid collected was *Javesella pellucida* (Fabricius, 1794), a generalist on grasses with a holarctic distribution. This well-studied species is a pest of cultivated grasses, mainly as a vector of plant diseases (Mochida and Kisimoto, 1971). The usefulness of monitoring this species on the KENWR is not apparent.

Culicidae As larvae, mosquitoes are filter feeders in a variety of fresh and brackish water bodies as well as small pools and containers. Most adults feed on blood of vertebrates. Both larvae and adults are prey for invertebrates and vertebrates.

Adult Culicidae can be reliably identified to species. Unfortunately the LTEMP methods, particularly the roughness of sweep net sampling and storage in alcohol, damaged most Culicidae specimens so that they could not be identified with confidence. As long as these methods are to be used, Culicidae could only be monitored at the family resolution. Relative abundance of mosquitoes over time may have monitoring value since they are prey for birds and bats.

Ichneumonidae Like the Braconidae, Ichneumonids are parasitoids of diverse insects. In the LTEMP samples, Ichneumonidae appeared to be represented by many species, each infrequently collected. Unfortunately, identification of most members of this large group is

not feasible at present given the available expertise and literature. As with the Braconidae, monitoring of Ichneumonidae at the family resolution is not justified, and probably no single Ichneumonid species could be monitored well under LTEMP methods.

Muscidae The Muscids (including Fanniinae) often feed on decaying organic matter such as dung or rotting plants. The Muscidae in the LTEMP samples appear to be represented by quite a few species, each infrequently collected, but one species tentatively identified as *Fannia spathiophora* appears to be fairly common. Muscids were often collected in large numbers where present. Although good keys for this group exist (e.g., Chillcott, 1960; Huckett, 1965; Huckett and Vockeroth, 1987), determinations were difficult for me in the absence of a reference collection. I have not been able to obtain help with identifications of specimens in this group.

Psyllidae The Psyllids feed on juices of plants. The two common genera encountered in sweep net samples, *Craspedolepta* and *Psylla* are, respectively, monophagous on fireweed, *Chamerion angustifolium*, or polyphagous on various species of willows, *Salix* spp. Identifications of adult Psyllids to species were generally possible using available keys (e.g., Hodkinson, 1978) and through the help of Dr. Robert Foottit.

Simuliidae Larvae of Simuliidae are filter feeders in flowing waters; the adults feed on the blood of vertebrates. As with the Culicidae, Simuliids were attracted to the collectors. The genera *Simulium* and *Prosimulium* have been identified from LTEMP sweep net samples. A good key exists for species identifications (Adler et al., 2004), but I have not attempted to key the species yet.

Sminthuridae Smithurids may feed on vascular plants (Hopkin, 1997), fungi (Borror et al., 1989), or pollen (Kevan and Kevan, 1970). Possibly all of the Sminthuridae collected in LTEMP sweep net samples comprise two species: *Sminthurus* sp. A and a *Ptenothrix* sp., although another species, *Pseudobourletiella spinata* (MacGillivray, 1893), has also been collected on the KENWR in a sweep net sample in a separate project. The vast majority of the Smithuridae present in the samples were determined to be Sminthurus sp. A by Dr. Richard Snider (Michigan State University) (see Christiansen and Bellinger, 1998, for a key

including this species). At least some species of *Sminthurus* feed on the tissues of living vascular plants. I have collected *Sminthurus* sp. A on the dwarf shrubs *Ledum palustre* L. and *Empetrum nigrum* L., but I have not been able to observe feeding. One species in this genus, *Sminthurus viridis*, is an important pest of clover in Australia (Hopkin, 1997). Species of *Ptenothrix* are often found on fungi (Borror et al., 1989). On KENWR, I have collected *Ptenothrix* on lichens and on sporocarps of Russulaceae and Boletaceae. I have witnessed feeding of *Ptenothrix* in the spore bearing tissues of a bolete.

The Sminthuridae present in the LTEMP samples can easily be sorted to family and species, the three species so far collected on the KENWR being conspicuously morphologically distinct. The two species present in the sweep net samples are as yet undescribed, but Richard Snider is presently working on a revision of this group (personal communication). The *Ptenothrix* species may not be common enough (collected in about 10% of samples) to be used as a monitoring metric, but *Sminthurus* sp. A, comprising almost all of the Sminthuridae collected, could be monitored better than most families.

Even though they could be monitored with reasonable power, the value of Smithurids species as monitoring metrics is not clear because biological interactions of Sminthuridae are less well known than those of many other groups. Because of the abundance and ubiquity of Smithuridae over the landscape of the KENWR, investigation into the biology of the species, particularly their feeding preferences, is warranted.

Selection of taxa to be monitored As more species are identified from among the specimens collected on LTEMP, many additional species will be recognized as having high enough detection probabilities to serve as potential monitoring metrics. In 2007, KENWR staff collected additional samples, as yet unprocessed, using spatial sub-sampling that will allow estimation of detection probabilities. In addition, multiple surveys were conducted over the season. Analysis of these samples will enable estimation of detection probabilities of all arthropod taxa. To further broaden the scope of the taxa considered, identifications may not be limited to classical morphological taxonomy. KENWR's managers will consider employing molecular methods such as DNA barcoding (e.g., DeSalle and Amato, 2004; Ratnasingham and Hebert, 2007; Hajibabaei et al., 2007) so that even undescribed species may be recognized.

Even though particular arthropod species identified from LTEMP sweep net samples may not appear to warrant particular attention on their own, it is of great interest to the managers of the KENWR to monitor for the up-slope distribution shifts that are expected in coming decades. The strength of the multi-species approach increases as more taxa are included, so whole communities can be considered. Monitoring of arthropods, organisms that may respond quickly to climatic stimuli through dispersal and differential success, may prove to be particularly relevant in the context of climate change.

Given the limited amount of time and resources which could be spent on sorting and identification of arthropods and that no taxon seemed much more useful than others as a monitoring metric, I recommend that the species which are collected by sweep net with high probability of detection and which can be most easily identified be considered for monitoring. Based on my present experience, species of the families Delphacidae and Sminthuridae generally meet these criteria. As identifications continue and as detection probabilities become estimable, this list should grow considerably.

Conclusions

As proposed in 2004, the LTEMP sampling program served as an initial inventory of diverse biota and provided ample spatial data for modelling species distributions. By accounting for imperfect detection in the future, the LTEMP program will be poised to monitor change in the distribution of species over time on the landscape of the KENWR (at least for species with detection probability ≥ 0.5). This will allow managers of the KENWR to recognize shifts in the distributions of species on the KENWR. Of the two proposed sampling designs, I recommended the rotating panel design because it would provide more timely information about distributions, local colonization, and local extinction. Although monitoring for distribution shifts was mostly discussed in the context of species' response to climate, these same monitoring methods should serve well for monitoring the spread of exotic species.

References

Adler, P. H., D. C. Currie, and D. M. Wood. 2004. The black blies (Simuliidae) of North America. Cornell University Press, Ithaca, New York.

Alaska Department of Fish and Game. 2006. Our wealth maintained: a strategy for conserving Alaska's diverse wildlife and fish resources. Alaska Department of Fish and Game, Juneau, Alaska.

Ashworth, A. C., 2001. Chapter 8: Perspectives on Quaternary beetles and climate change. Pages 153–168 *in* Geological Perspectives of Global Climate Change. American Association of Petroleum Geologists Studies in Geology #47, Tulsa, Oklahoma. URL `http://www.ndsu.nodak.edu/instruct/ashworth/aapg_perspectives.pdf`.

Bailey, R. C., R. H. Norris, and T. B. Reynoldson. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. Journal of the North American Benthological Society **20**:280–286.

Bailey, S.-A., C. Horner-Devine, G. Luck, L. A. Moore, K. M. Carney, S. Anderson, C. Betrus, and E. Fleishman. 2004. Primary productivity and species richness: relationships among functional guilds, residency groups and vagility classes at multiple spatial scales. Ecography **27**:207–217.

Balmford, A., A. H. M. Jayasuriya, and M. J. B. Green. 1996. Using higher-taxon richness as a surrogate for species richness: II. Local applications. Proceedings of the Royal Society of London B **263**:1571–1575.

Barry, R. G. 1992. Mountain weather and climate. Routledge, London.

Bates, H. W. 1862. Contributions to an insect fauna of the Amazon Valley. Lepidoptera: Heliconidae. Transactions of the Linnean Society **23**:495–566.

Beaver, R. A. 1979. Host specificity of temperate and tropical animals. Nature (Lond.) **281**:139–141.

Bechtel, T. J., and B. J. Copeland. 1970. Fish species diversity indices as indicators of pollution in Galveston Bay, Texas. Contributions in Marine Science **15**:103–32.

Beirne, B. P. 1956. Leafhoppers (Homoptera: Cicadellidae) of Canada and Alaska. The Canadian Entomologist **88, supplement 2**:1–180.

Benin, S., B. Gebremedhin, M. Smale, J. Pender, and S. Ehui, 2003. Determinants of Cereal Diversity in Communities and on Household Farms of the Northern Ethiopian Highlands. Working Papers 03-14, Agricultural and Development Economics Division of the Food and Agriculture Organization of the United Nations (FAO - ESA). URL `http://ideas.repec.org/p/fao/wpaper/0314.html`.

Berg, E. E. 2000. Studies in the wilderness areas of the Kenai National Wildlife Refuge: fire, bark beetles, human development, and climate change. USDA Forest Service Proceedings **3**:63–67. URL `http://www.fs.fed.us/rm/pubs/rmrs_p015_3/rmrs_p015_3_063_067.pdf`.

Bivand, R., with contributions by Luc Anselin, O. Berke, A. Bernat, M. Carvalho, Y. Chun, C. Dormann, S. Dray, R. Halbersma, N. Lewin-Koh, G. Millo, W. Mueller, H. Ono, P. Peres-Neto, M. Reder, M. Tiefelsdorf, , and D. Yu., 2007. `spdep`: Spatial dependence: weighting schemes, statistics and models.

Bohlen, P. J., P. M. Groffman, T. J. Fahey, M. C. Fisk, E. Suarez, D. M. Pelletier, and R. T. Fahey. 2004. Ecosystem consequences of exotic earthworm invasion of north temperate forests. Ecosystems **7**:1–12.

Boos, D. D., A. D. Brooks, and D. Nychka, 2006. Rlab: Functions and Datasets Required for ST370 class. URL `http://www.r-project.org,http://courses.ncsu.edu/ST370/distance/rlab`.

Borror, D. J., C. A. Triplehorn, and N. F. Johnson. 1989. An introduction to the study of insects. Sixth edition. Harcourt Brace College Publishers, New York.

Borror, D. J., and R. E. White. 1970. A field guide to insects: America north of Mexico. Houghton Mifflin Company, New York.

Boulinier, T., J. D. Nichols, J. R. Sauer, J. E. Hines, and K. H. Pollock. 1998. Estimating species richness: the importance of heterogeneity in species detectability. Ecology **79**:1018–1028.

Bowman, M. F., and R. C. Bailey. 1997. Does taxonomic resolution affect the multivariate description of the structure of freshwater benthic macroinvertebrate communities? Canadian Journal of Fisheries and Aquatic Sciences **54**:1802–1807.

Box, E. O., B. N. Holben, and V. Kalb. 1989. Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net CO_2 flux. Vegetatio **80**:71–89.

Breiman, L. 2001. Random Forests. Machine Learning **45**:5–32.

Breiman, L., 2002. Manual on setting up, using, and understanding random forests v3.1. URL `http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf`.

Brown, J. H., and D. W. Davidson. 1977. Competition between seed-eating rodents and ants in desert ecosystems. Science **196**:880–882.

Busby, J. R., 1991. BIOCLIM - a bioclimatic analysis and prediction system. Pages 64–68 *in* C. R. Margules and M. P. Austin, editors. Nature Conservation: cost effective biological surveys and data analysis. CSIRO, Canberra, Australia.

Cameron, A. C., and F. A. G. Windmeijer. 1996. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics **14**:209–20.

Chapin, S. F., III, E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U. Hooper, S. Lavorel, O. E. Sala, S. E. Hobbie, M. C. Mack, and S. Díaz. 2000. Consequences of changing biodiversity. Nature **405**:234–242.

Cheng, L., editor. 1976. Marine Insects. American Elsevier Publishing Company, New York.

Chillcott, J. G. 1960. A revision of the Nearctic species of Fanninae (Diptera: Muscidae). The Canadian Entomologist **92, supplement 14**:1–295.

Christiansen, K. A., and P. Bellinger. 1998. The Collembola of North America North of the Rio Grande. Second edition. Grinnell College, Grinnell, Iowa.

Collins, N. M., 1989. Termites. Pages 455–471 *in* H. Lieth and M. J. A. Werger, editors. Ecosystems of the World. Vol. 14B. Tropical rain forest ecosystems: Biogegraphical and ecological studies. Elsevier, Amsterdam.

Currie, D. J. 1991. Energy and large-scale patterns of animal- and plant-species richness. The American Naturalist **137**:27–49.

Danks, H. V., 1979. Canada and its insect fauna: an overview. Pages 567–573 *in* H. V. Danks, editor. Canada and its insect fauna. Memoirs of the Entomological Society of Canada 108. The Entomological Society of Canada, Ottawa, Ontario, Canada.

Danks, H. V., 1988. Insects of Canada. Document Series 1, Biological Survey of Canada (Terrestrial Arthropods).

Danks, H. V., 1996. How to assess insect biodiversity without wasting your time. Technical Report 5, Biological Survey of Canada (Terrestiral Arthropods), Ottawa, Ontario, Canada.

Danks, H. V., J. A. Downes, D. J. Larson, and G. G. E. Scudder, 1997. Insects of the Yukon: characteristics and history. Pages 963–1013 *in* H. V. Danks and J. A. Downes, editors. Insects of the Yukon. Biological Survey of Canada, Ottawa, Ontario, Canada.

Danks, H. V., and R. G. Foottit. 1989. Insects of the boreal zone of Canada. The Canadian Entomologist **121**:625–690.

DeSalle, R., and G. Amato. 2004. The expansion of conservation genetics. Nature Reviews Genetics **5**:702–712.

Dixon, A. F. G. 1977. Aphid Ecology: Life Cycles, Polymorphism, and Population Regulation. Annual Review of Ecology and Systematics **8**:329–353.

Dixon, A. F. G. 1985. Aphid Ecology. Blackie & Son Limited, New York.

Dixon, A. F. G., P. Kindlmann, and J. Holman. 1987. Why are there so few species of aphids, especially in the tropics. American Naturalist **129**:580–592.

Dorazio, R. M., and J. A. Royle. 2005. Estimating size and composition of biological communities by modeling the occurrence of species. Journal of the American Statistical Association **100**:389–398.

Egloff, D. A., and W. H. Brakel. 1973. Stream pollution and a simplified diversity index. Journal Water Pollution Control Federation **45**:2269–2275.

Fabricius, J. C. 1794. Entomologicae Systematica.

Fernandes, G. W., and P. W. Price. 1988. Biogeographical gradients in galling species richness. Oecologia **76**:161–167.

Finnamore, A. T. 1996. The advantages of using arthropods in ecosystem management. Biological Survey of Canada (Terrestrial Arthropods), Ottawa, Ontario.

Fisher, A. G. 1960. Latitudinal variations in organic diversity. Evolution **14**:64–81.

Fleishman, E., G. T. Austin, and A. Weiss. 1988. An empirical test of Rapoport's rule: elevational gradients in montane butterfly communities. Ecology **79**:2472–2483.

Fleishman, E., J. P. Fay, and D. D. Murphy. 2000. Upsides and downsides: contrasting topographic gradients in species richness and associated scenarios for climate change. Journal of Biogeography **27**:1209–1219.

Frelich, L. E., C. M. Hale, S. Scheu, A. R. Holdsworth, L. Heneghan, P. J. Bohlen, and P. B. Reich. 2006. Earthworm invasion into previously earthworm-free temperate and boreal forests. Biological Invasions **8**:1235–1245.

Gaston, K. J. 1992. Regional numbers of insect and plant species. Functional Ecology **6**:243–247.

Gauld, I. D. 1987. The so-called anomalous diversity of tropical Ichneumonidae re-examined. Biological Journal of the Linnean Society **30**:299–312.

Gelfand, A. E., J. A. Silander, Jr., A. L. S. Wu, P. O. Lewis, A. G. Rebelo, and M. Holder. 2003. Explaining Species Distribution Patterns through Hierarchical Modeling. Bayesian Analysis **1**:1–35.

Gottfried, M., H. Pauli, K. Reiter, and G. Grabherr. 1999. A fine-scaled predictive model for changes in species distribution patterns of high mountain plants induced by climate warming. Diversity and Distributions **5**:241–251.

Goulet, H., and J. T. Huber, editors. 1993. Hymenoptera of the world : an identification guide to families. Centre for Land and Biological Resources Research, Ottawa, Ontario, Canada.

Grime, J. P. 1973*a*. Competitive exclusion in herbaceous vegetation. Nature **242**:344–347.

Grime, J. P. 1973*b*. Control of species density on herbaceous vegetation. Journal of Environmental Management **1**:51–67.

Groves, C. R. 2003. Drafting a conservation blueprint: a practitioner's guide to planning for biodiversity. Island Press, Washington.

Grytnes, J. A., and C. M. McCain. 2007. Elevational trends in biodiversity. Encylopedia of Diversity **2**:1–8.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8**:993–1009.

Hajibabaei, M., G. A. C. Singer, E. L. Clare, and P. D. N. Hebert. 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. BMC Biology **5**:24.

Hannah, L., G. Midgley, and D. Millar. 2002. Climate change-integrated conservation strategies. Global Ecology & Biogeography **11**:485–495. URL `http://ecology.botany.ufl.edu/adveco/downloads/Hannah_2002_Climate%20change-integrated%20conservation%20strategies.pdf`.

Heino, J., T. Muotka, R. Paavola, and L. Paasivirta. 2003. Among-taxon congruence in biodiversity patterns: can stream insect diversity be predicted using single taxonomic groups? Canadian Journal of Fisheries and Aquatic Sciences **60**:1039–1049.

Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. American Naturalist **163**:192–211.

Hines, J., 2007. Program PRESENCE. USGS Patuxent Wildlife Research Center, Laurel, Maryland. URL `http://www.mbr-pwrc.usgs.gov/software/presence.html`.

Hodkinson, I. D. 1978. The psyllids (Homoptera: Psylloidea) of Alaska. Systematic Entomology **3**:333–360.

Hooper, D. U., S. F. Chapin, III, J. J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J. H. Lawton, D. M. Lodge, M. Loreau, S. Naeem, B. Schmid, H. Stetälä, A. J. Symstad, J. Vandermeer, and D. A. Wardle. 2005. Effects of biodiversity on ecosystem funcitoning: a consensus of current knowledge. Ecological Monographs **75**:3–35.

Hopkin, S. P. 1997. Biology of the Springtails (Insecta: Collembola). Oxford University Press.

Hovanitz, W., 1958. Distribution of butterflies in the New World. Pages 321–368 *in* C. L. Hubbs, editor. Zoogeography. American Association for the Advancement of Science Publication 51.

Huckett, H. C. 1965. The Muscidae of northern Canada, Alaska and Greenland (Diptera). Memoirs of the Entomological Society of Canada **42**:1–369.

Huckett, H. C., and J. R. Vockeroth, 1987. Family Muscidae. Pages 1115–1131 *in* J. F. McAlpine, B. V. Peterson, G. E. Shewell, H. J. Teskey, J. R. Vockeroth, and D. M. Wood, editors. Manual of Nearctic Diptera, volume 2. Agriculture Canada Research Branch.

Huston, M. A. 1979. A general hypothesis of species diversity. American Naturalist **113**:81–101.

Jansen, E., J. Overpeck, K. Briffa, J. C. Duplessy, F. Joos, V. Masson-Delmotte, D. Olago, B. Otto-Bliesner, W. Peltier, D. R. O. S. R. V. S. Rahmstorf R. Ramesh, D. Raynaud, and D. Zhang, 2007. Palaeoclimate. Chapter 6, pages 433–497 *in* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, editors. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York. URL `http://ipcc-wg1.ucar.edu/wg1/Report/AR4WG1_Print_Ch06.pdf`.

Janzen, D. H. 1973. Sweep net samples of tropical foliage insects: effects of seasons, vegetation types, elevation, time of day, and insularity. Ecology **54**:687–708.

Karr, J. R., and D. N. Kimberling. 2003. A terrestrial arthropod index of biological integrity for shrubb-steppe landscapes. Northwest Science **77**:202–213.

Kaspari, M., S. O'Donnell, and J. R. Kercher. 2000. Energy, density, and constraints to species richness: ant assemblages along a productivity gradient. The American Naturalist **155**:280–293.

Kearns, C. A. 1992. Anthophilous fly distribution across an elevation gradient. American Midland Naturalist **127**:172–182.

Kempton, R. A. 1979. The structure of species abundance and measurement of diversity. Biometrics **35**:307–321.

Kennedy, C. H. 1928. Evolutionary level in relation to geographic, seasonal and diurnal distribution of insects. Ecology **9**:367–379.

Kevan, P. G., and D. K. M. Kevan. 1970. Collembola as pollen feeders and flower visitors with observations from the high arctic. Quaestiones entomologicae **6**:311–326.

Kouki, J., P. Niemelä, and M. Viitasaari. 1994. Reversed latitudinal gradient in species richness of sawflies (Hymenoptera, Symphyta). Annales Zoologici Fennici **31**:83–88.

Kremen, C., R. K. Colwell, T. L. Erwin, D. D. Murphy, R. F. Noss, and M. A. Sanjayan. 1993. Terrestrial arthropod assemblages: their use in conservation planning. Conservation Biology **7**:796–808.

Kusnezov, M. 1957. Numbers of species of ants in faunae of different latitudes. Evolution **11**:298–299.

Lawton, J. H. 1999. Are there general laws in ecology? Oikos **84**:177–192.

Leibold, M. A. 1996. A graphical model of keystone predators in food webs: trophic regulation of abundance, incidence, and diversity patterns in communities. The American Naturalist **147**:784–812.

Lenat, D. R., and V. H. Resh. 2001. Taxonomy and stream ecology—the benefits of genus- and species-level identifications. Journal of the North American Benthological Society **20**:287–298.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News **2**:18–22. URL `http://CRAN.R-project.org/doc/Rnews/`.

Lo Seen Chong, D., E. Mougin, and J. P. Gastellu-Etchegorry. 1993. Relating the global vegetation index to net primary productivity and actual evapotranspiration over Africa. International Journal of Remote Sensing **14**:1517–1546.

Lomolino, M. V. 2001. Elevation gradients of species-density: historical and prospective views. Global Ecology & Biogeography **10**:3–13.

Losey, J. E., and M. Vaughan. 2006. The economic value of ecological services provided by insects. BioScience **56**:311–323.

Mac Nally, R., E. Fleishman, J. P. Fay, and D. D. Murphy. 2003. Modelling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the Great Basin of western North America. Biological Conservation **110**:21–31.

MacArthur, R. H. 1972. Geographical ecology; patterns in the distribution of species. Harper & Row, New York.

MacGillivray, A. D. 1893. North American Thysanura. The Canadian Entomologist **25**:313–318.

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology **84**:2200–2207.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling. Elsevier, New York.

MacKenzie, D. I., and J. A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology **42**:1105–1114.

Magness, D. R., F. Huettmann, and J. M. Morton. 2008. Using random forests to provide predicted species distribution maps as a metric for ecological inventory & monitoring programs. Studies in Computational Intelligence **122**:209–229.

Magurran, A. E. 1988. Ecological diversity and its measurement. Princeton University Press, Princeton, New Jersey.

Mandelik, Y., T. Dayan, V. Chikatunov, and V. Kravchenko. 2007. Reliability of a Higher-Taxon Approach to Richness, Rarity, and Composition Assessments at the Local Scale. Conservation Biology **21**:1506–1515.

Marshall, J. C., A. L. Steward, and B. D. Harch. 2006. Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an Australian dryland river: the benefits and costs of using species abundance data. Hydrobiologia **572**:171–194.

McAlpine, J. F., B. V. Peterson, G. E. Shewell, H. J. Teskey, J. R. Vockeroth, and D. M. Wood, editors. 1981. Manual of Nearctic Diptera, volume 1. Research Branch, Agriculture Canada, Ottawa, Ontario, Canada.

McAlpine, J. F., B. V. Peterson, G. E. Shewell, H. J. Teskey, J. R. Vockeroth, and D. M. Wood, editors. 1987. Manual of Nearctic Diptera, volume 2. Research Branch, Agriculture Canada, Ottawa, Ontario, Canada.

McCain, C. M. 2005. Elevational gradients in diversity of small mammals. Ecology **86**:366–372.

McCain, C. M. 2007. Could temperature and water availability drive elevational species richness patterns? A global case study for bats. Global Ecology and Biogeography **16**:1–13.

McCann, K. S. 2000. The diversity-stability debate. Nature **405**:228–233.

McCoy, E. D. 1990. The distribution of insects along elevational gradients. Oikos **58**:313–332.

Michener, C. D. 1979. Biogeography of the bees. Annals of the Missouri Botanical Garden **66**:277–347.

Migge-Kleian, S., M. A. McLean, J. C. Maerz, and L. Heneghan. 2006. The influence of invasive earthworms on indigenous fauna in ecosystems previously uninhabited by earthworms. Biological Invasions **8**:1275–1285.

Mittelbach, G. G., D. W. Schemske, H. V. Cornell, A. P. Allen, J. M. Brown, M. B. Bush, S. P. Harrison, A. H. Hurlbert, N. Knowlton, H. A. Lessios, C. M. McCain, A. R. McCune, L. A. McDade, M. A. McPeek, T. J. Near, T. D. Price, R. E. Ricklefs, K. Roy, D. F. Sax, D. Schluter, J. M. Sobel, and M. Turelli. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecology Letters **10**:315–331.

Mittelbach, G. G., C. F. Steiner, S. M. Scheiner, K. L. Gross, H. L. Reynolds, R. B. Waide, M. R. Willig, S. I. Dodson, and L. Gough. 2001. What is the observed relationship between species richness and productivity? Ecology **82**:2381–2396.

Mochida, O., and R. Kisimoto. 1971. A review of the studies of *Javesella pellucida* (F.) (Hom., Delphacidae) and associated subjects. Review of Plant Protection Research **4**:1–57.

Morton, S. R., and D. W. Davidson. 1988. Comparative structure of harvester ant communities in arid Australia and North America. Ecological Monographs **58**:19–38.

Murdoch, W. W., F. C. Evans, and C. H. Peterson. 1972. Diversity and patterns in plants and insects. Ecology **53**:819–829.

Myneni, R. B., F. G. Hall, P. J. Sellers, and A. L. Marshak. 1995. The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience and Remote Sensing **33**:481–486.

Myneni, R. B., C. D. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani. 1997. Increased plant growth in the northern high latitudes from 1981 to 1991. Nature **386**:698–702.

O'Brien, L., 2006. Kenai Peninsula Land Cover Classification. Kenai National Wildlife Refuge, Soldotna, Alaska.

Oksanen, J., R. Kindt, P. Legendre, and R. B. O'Hara, 2007. `vegan`: Community Ecology Package. URL `http://cran.r-project.org/`.

Oliver, I., and A. J. Beattie. 1996. Designing a cost-effective invertebrate survey: a test of methods for rapid assessment of biodiversity. Ecological Applications **6**:594–607.

Olson, D. M. 1994. The distribution of leaf litter invertebrates along a neotropical altitudinal transect. Journal of Tropical Ecology **10**:129–150.

Paoletti, M. G., and M. Bressan. 1996. Soil invertebrates of bioindicators of human disturbance. Critical Revies in Plant Sciences **15**:21–62.

Parmesan, C. 1996. Climate and species' range. Nature **382**:765–766.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. Annual Review of Ecology and Systematics **37**:637–669.

Parmesan, C., N. Ryrholm, C. Stefanescu, J. K. Hill, C. D. Thomas, H. Descimon, B. Huntley, L. Kaila, J. Kullberg, T. Tammaru, W. J. Tennent, J. A. Thomas, and M. Warren. 1999. Poleward shifts in geographic ranges of butterfly species associated with regional warming. Nature **399**:579–583.

Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature **421**:37–42.

Pianka, E. R. 1966. Latitudinal gradients in species diversity: a review of concepts. The American Naturalist **100**:33–46.

Pik, A. J., I. Oliver, and A. J. Beattie. 1999. Taxonomic sufficiency in ecological studies of terrestrial invertebrates. Australian Journal of Ecology **24**:555–562.

Prendergast, J. R., R. M. Quinn, J. H. Lawton, B. C. Eversham, and D. W. Gibbons. 1993. Rare species, the coincincidence of diversity hotspots and conservation strategies. Nature **365**:335–337.

Price, P. W., G. W. Fernandez, A. C. F. Lara, J. Brawn, H. Barrios, M. G. Wright, S. P. Ribeiro, and N. Rothcliff. 1998. Global patterns in local number of insect galling species. Journal of Biogeography **25**:581–591.

Quicke, D. L. J., and R. A. Kruft. 1995. Latitudinal gradients in North American braconid wasp species richness and biology. Journal of Hymenoptera Research **4**:194–203.

Raftery, A., J. Hoeting, C. Volinsky, I. Painter, and K. Y. Yeung, 2006. The BMA Package. URL `http://www.r-project.org,http://www.research.att.com/~volinsky/bma.html`.

Rahbek, C. 1995. The elevational gradient of species richness: a uniform pattern? Ecography **18**:200–205.

Rahbek, C. 2005. The role of spatial scale and the perception of large-scale species-richness patterns. Ecology Letters **8**:224–239.

Rapoport, E. H. 1975. Areografía: estrategias geográficas de las especies. Fondo de Cultura Económica, Mexico City.

Ratnasingham, S., and P. D. N. Hebert. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes **7**:1–10. URL `http://www.barcodinglife.org/libhtml/docs/bold.pdf`.

R Development Core Team, 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`.

Ribeiro, P. J., Jr., and P. J. Diggle. 2001. `geoR`: a package for geostatistical analysis. R News **1**:14–18. URL `http://CRAN.R-project.org/doc/Rnews/`.

Ricklefs, R. E., and K. O'Rourke. 1975. Aspect diversity in moths: a temperate-tropical comparison. Evolution **29**:313–324.

Ricotta, C., and G. C. Avena. 1998. Fractal modelling of the remotely sensed two-dimensional net primary production pattern with annual cumulative AVHRR NDVI data. Remote Sensing **19**:2413–2418.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, New York.

Roth, D. S., I. Perfecto, and B. Rathcke. 1994. The effects of management systems on ground-foraging ant diversity in Costa Rica. Ecological Applications **4**:423–436.

Royle, J. A., and M. Kéry. 2007. A Bayesian state-space formulation of dynamic occupancy models. Ecology **88**:1813–1823.

Royle, J. A., M. Kéry, R. Guatier, and H. Schmid. 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. Ecological Monographs **77**:465–481.

Royle, J. A., and W. A. Link. 2006. Generalized site occupancy models allowing for false positive and false negative errors. Ecology **87**:835–841.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology **84**:777–790.

Sanchez-Rodriguez, J. F., and A. Baz. 1995. The effects of elevation on the butterfly communities of a Mediterranean mountain, Sierra de Javalmbre, central Spain. Journal of the Lepidopterists' Society **49**:192–207.

Sanders, N. J. 2002. Elevational gradients in ant species richness: area, geometry, and Rapoport's rule. Ecography **25**:25–32.

Schreiber, H. 1978. Dispersal centres of Sphingidae (Lepidoptera) in the Neotropical region. Biogeographica 10. Junk, The Hague.

Scriber, J. M. 1973. Latitudinal gradients in larval feeding specialization of the world Papilionidae (Lepidoptera). Psyche **80**:355–373.

Scriber, J. M. 1984. Larval foodplant utilization by the world Papilionidae (Lep.): latitudinal gradients reappraised. Tokuruna (Acta Rhopalocerologica) **6**/**7**:1–50.

Seimann, E. 1998. Experimental tests of effects of plant productivity and diversity on grassland arthropod diversity. Ecology **79**:2057–2070.

Sime, K. R., and A. V. Z. Brower. 1998. Explaining the latitudinal gradient anomoly in ichneumonid species richness: evidence from butterflies. Journal of Animal Ecology **67**:387–399.

Simpson, J. J., G. L. Hufford, C. Daly, J. S. Berg, and M. D. Fleming. 2005. Comparing maps of mean monthly surface temperature and precipitation for Alaska and adjacent areas of Canada produced by two different methods. Arctic **58**:137–161. URL `http://repositories.cdlib.org/cgi/viewcontent.cgi?article=2934&context=postprints`.

Slansky, F., Jr. 1973. Latitudinal gradients in species diversity of the New World swallowtail butterflies. Journal of Research on the Lepidoptera **11**:201–217.

Soudani, K., C. François, G. le Maire, V. Dantec, and E. Dufrêne. 2006. Comparative analysis of IKONOS, SPOT, and ETM+ data for leaf area index estimation in temperate coniferous and deciduous forest stands. Remote Sensing of Environment **102**:161–175.

Sparrow, H. R., T. D. Sisk, P. R. Ehrlich, and D. D. Murphy. 1994. Techniques and guidelines for monitoring neotropical butterflies. Conservation Biology **8**:800–809.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn, 2003. WinBUGS User Manual Version 1.4. URL `http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf`.

Srivastava, D. S., and J. H. Lawton. 1998. Why more productive sites have more species: an experimental test of theory using tree-hole communities. The American Naturalist **152**:1998.

Stanton, N. L. 1979. Patterns of species diversity in temperate and tropical litter mites. Ecology **60**:295–304.

Stevens, G. C. 1989. The latitudinal gradient in geographical range: how so many species coexist in the tropics. The latitudinal gradient in geographical range: how so many species coexist in the tropics **133**:240–256.

Stevens, G. C. 1992. The elevational gradient in altitudinal range: an extension of Rapoport's laitudinal rule to altitude. American Naturalist **140**:893–911.

Taylor, L. R., 1978. Bates, Williams, Hutchinson — a variety of diversities. Pages 1–18 *in* L. A. Mound and N. Warloff, editors. Diversity of Insect Faunas: 9th Symposium of the Royal Entomological Society. Oxford.

Teraguchi, S. J., J. Stenzel, J. Sedlacek, and R. Deininger. 1981. Arthropod-grass communities: comparison of communities in Ohio and Alaska. Journal of Biogeography **8**:53–65.

Thomas, A. 1994. BUGS: a statistical modelling package. RTA/BCS Modular Languages Newsletter **2**:36–38.

Thomas, R. H., and S. F. MacLean, Jr. 1988. Community structure in soil Acari along a latitudinal transect of tundra sites in northern Alaska. Pedobiologia **31**:113–138.

Tillyard, R. J. 1917. The biology of dragonflies. Cambridge University Press, Cambridge.

Tilman, D. 1982. Resource competition and community structure. Princeton University Press, Princeton, New Jersey.

Turner, B. D., and E. Broadhead. 1974. Diversity and distribution of psocid populations on *Mangifera indica* L. in Jamaica and their relationships to altitude and microepiphyte diversity. Journal of Animal Ecology **43**:173–190.

USDA Forest Service. 2000-2003. Field procedures for the south-central Alaska inventory. Pacific Northwest Research Station, Anchorage.

Van Deusen, P. C. 1996. Incorporating predictions into an annual forest inventory. Canadian Journal of Forest Research **26**:1709–1713.

Van Deusen, P. C. 1999. Modeling trends with annual survey data. Canadian Journal of Forest Research **29**:1824–1828.

Van Deusen, P. C. 2002. Comparison of some annual forest inventory estimators. Canadian Journal of Forest Research **32**:1992–1995.

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. Fourth edition. Springer, New York. URL `http://www.stats.ox.ac.uk/pub/MASS4`.

Waite, I. R., A. T. Herlihy, D. P. Larsen, N. S. Urquhart, and D. J. Klemm. 2004. The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: an example from the Mid-Atlantic Highlands, U.S.A. Freshwater Biology **49**:474–489.

Wallace, A. R. 1853. On the habits of the butterflies of the Amazon Valley. Transactions of the Entomological Society of London **2**:253–264.

Wallace, A. R. 1878. Tropical Nature and Other Essays. Macmillan, London.

Walther, G., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J. Fromentin, O. Hoegh-Guldberg, and F. Bairlein. 2002. Ecological responses to recent climate change. Nature **416**:389–395.

Werner, R. A. 1983. Biomass, density, and nutrient content of plant arthropods in the taiga of Alaska. Canadian Journal of Forest Research **13**:729–739.

Wettstein, W., and B. Schmid. 1999. Conservation of Arthropod Diversity in Montane Wetlands: Effect of Altitude, Habitat Quality and Habitat Fragmentation on Butterflies and Grasshoppers. Journal of Applied Ecology **36**:363–373.

White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study **46**:120–138. URL `http://welcome.warnercnr.colostate.edu/~gwhite/mark/euring.PDF`.

White, R. E. 1983. A field guide to the beetles of North America. Houghton Mifflin Company, New York.

Whittaker, R. J., K. J. Willis, and R. Field. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of biogeography **28**:453–470.

Williams, P., L. Hannah, S. Andelman, G. Midgley, M. Araújo, G. Hughes, L. Manne, E. Martinez-Meyer, and R. Pearson. 2005. Planning for climate change: identifying minimum-dispersal corridors for the Cape Proteaceae. Conservation Biology **19**:1063–1074.

Wilson, E. O. 1987. The little things that run the world (the importance and conservation of invertebrates). Conservation Biology **1**:344–346.

Wilson, R. J., D. Gutiérrez, J. Gutiérrez, D. Martínez, R. Agudo, and V. J. Monserrat. 2005. Changes to the elevational limits and extent of species ranges associated with climate change. Ecology Letters **8**:1138–1146.

Wilson, S. W. 1988. Delphacidae of Alaska (Homoptera: Fulgoroidea). Great Basin Naturalist Memoirs **12**:335–343.

Wilson, S. W., 1997. Dephacid planthoppers of the Yukon. Pages 377–385 *in* H. V. Danks and J. A. Downes, editors. Insects of the Yukon. Biological Survey of Canada, Ottawa, Ontario, Canada.

Wintle, B. A., M. A. McCarthy, C. T. Volinsky, and R. P. Kavanagh. 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology **17**:1579–1590.

Wolda, H. 1987. Altitude, habitat and tropical insect diversity. Biological Journal of the Linnean Society **30**:313–323.

Wright, D. H. 1983. Species-energy: an extension of species-area theory. Oikos **41**:496–506.

Wu, R. S. S. 1982. Periodic defaunation and recovery in a subtropical epibenthic community, in relation to organic pollution. Journal of experimental marine biology and ecology **64**:253–269.

Zeileis, A., C. Kleiber, and S. Jackman, 2007. Regression models for count data in R. Technical Report 53, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Wien, Austria. URL `http://epub.wu-wien.ac.at/`.

Appendix A.

Full results of Bayesian Model Averaging analyses

Table A.1: Full results of BMA analyses: family richness.

	BMA results			Single-variable regression results				
Coefficient	$p \neq 0$	mean	SD	mean	SE	p	R^2_{dev}	AIC
Intercept	100	-1.5E+00	9.2E-01	2.0E+00	4.2E-02	<2e-16	0.00	1456.80
S_{veg}	100	2.0E-02	4.9E-03	2.4E-02	6.1E-03	5.8E-05	0.05	1444.60
h_{alpine}	100	-5.6E-01	1.3E-01	-4.7E-01	1.2E-01	8.5E-05	0.05	1444.16
Day	100	1.9E-02	5.4E-03	2.1E-02	7.4E-03	4.7E-03	0.03	1450.90
h_{hemlock}	100	-6.2E-01	1.8E-01	-6.1E-01	2.2E-01	5.0E-03	0.03	1451.26
P_{A}	99.3	-1.4E-03	3.2E-04	-5.6E-04	8.6E-05	8.7E-11	0.13	1418.23
T_{O}:P_{A}	99.3	9.4E-05	1.5E-05	-8.6E-06	6.9E-06	2.1E-01	0.01	1457.32
$NDVI$:P_{A}	95.4	1.1E-03	4.1E-04	1.4E-03	1.8E-04	3.8E-14	0.15	1410.18
$h_{\text{halophytic wetland}}$	49.7	4.0E-01	4.8E-01	8.8E-01	6.1E-01	1.5E-01	0.01	1456.27
$NDVI$:E	16.5	-2.3E-04	5.7E-04	1.6E-03	3.2E-04	1.0E-06	0.06	1441.23
T_{A}:P_{A}	14	4.6E-05	1.5E-04	1.0E-04	7.2E-05	1.6E-01	0.01	1457.30
h_{snow}	12.8	-1.5E-01	4.9E-01	-2.9E+00	7.5E-01	1.1E-04	0.09	1431.47
E:P_{A}	10.7	4.9E-08	1.6E-07	-4.3E-07	6.5E-08	6.0E-11	0.14	1415.24
E:T_{A}	9.9	-5.5E-05	2.1E-04	-2.9E-04	9.9E-05	3.7E-03	0.03	1451.42
$h_{\text{mixed deciduous}}$	8.4	-5.2E-02	2.1E-01	-4.6E-01	5.2E-01	3.8E-01	0.00	1458.05
T_{O}:T_{A}	7.7	4.5E-04	1.8E-03	2.5E-02	3.1E-03	9.1E-16	0.17	1405.28
h_{alder}	6	1.2E-02	5.7E-02	2.0E-01	1.7E-01	2.3E-01	0.01	1457.30
T_{A}	5.8	7.2E-03	3.9E-02	3.2E-01	5.3E-02	2.3E-09	0.10	1427.35
$NDVI$:T_{A}	5.7	-3.3E-02	1.6E-01	8.3E-01	1.1E-01	2.0E-13	0.14	1415.67
$NDVI$	5.5	1.1E-01	5.1E-01	2.1E+00	2.3E-01	2.7E-20	0.22	1388.00
T_{O}:$Hour$	5	1.0E-04	5.4E-04	5.9E-03	9.8E-04	2.2E-09	0.11	1425.84
Day:E	2.7	9.3E-08	1.1E-06	-4.9E-06	6.8E-07	6.9E-13	0.15	1412.57
E	2.6	1.4E-05	1.8E-04	-8.6E-04	1.2E-04	1.7E-13	0.15	1410.29
$Hour$	2.6	7.2E-04	5.4E-03	4.0E-02	2.5E-02	1.1E-01	0.01	1456.21
h_{sparse}	2.3	1.2E-02	1.1E-01	-9.8E-01	4.8E-01	3.9E-02	0.01	1454.62
T_{O}:E	1.5	7.0E-07	8.6E-06	-3.6E-05	1.2E-05	1.8E-03	0.03	1448.77
$Area$	1.4	1.6E-06	1.7E-05	4.0E-04	7.0E-05	1.3E-08	0.09	1429.99

continued on next page...

Table A.1 continued.

Coefficient	$p \neq 0$	mean	SD	mean	SE	p	R^2_{dev}	AIC
h_{willow}	1.4	-2.9E-03	3.3E-02	5.9E-02	2.8E-01	8.3E-01	0.00	1458.76
$NDVI{:}T_{\text{O}}$	1.3	4.1E-04	5.0E-03	2.0E-01	1.8E-02	6.0E-29	0.30	1357.03
Sky	1.3	-5.1E-04	5.9E-03	-1.4E-01	4.3E-02	1.2E-03	0.03	1449.76
$h_{\text{alder-willow}}$	1.2	3.2E-03	4.2E-02	5.3E-01	4.4E-01	2.3E-01	0.01	1457.23
$h_{\text{cottonwood}}$	1.1	3.7E-03	5.2E-02	6.9E-01	6.2E-01	2.7E-01	0.00	1457.40
$Slope$	1	-2.1E-05	3.4E-04	-9.5E-03	2.7E-03	3.6E-04	0.05	1445.40
$h_{\text{white spruce}}$	1	6.9E-04	1.2E-02	2.8E-01	1.4E-01	4.3E-02	0.01	1454.62
humidity	1	2.0E-05	3.2E-04	5.6E-03	3.3E-03	9.4E-02	0.01	1455.99
$h_{\text{mixed forest}}$	0.9	3.6E-04	7.9E-03	3.3E-01	1.1E-01	2.0E-03	0.03	1449.16
$Day{:}T_{\text{O}}$	0.8	2.1E-07	1.0E-05	5.0E-04	6.5E-05	1.9E-14	0.16	1407.42
T_{O}	0.8	5.8E-05	1.8E-03	8.3E-02	1.1E-02	1.6E-13	0.15	1410.84
h_{barren}	0.8	2.1E-03	4.7E-02	-1.5E+00	4.3E-01	4.7E-04	0.04	1445.63
$h_{\text{shrub wetland}}$	0.8	3.2E-04	1.5E-02	1.4E-01	2.5E-01	5.7E-01	0.00	1458.47
$h_{\text{paper birch}}$	0.8	-3.8E-04	1.4E-02	1.3E-01	2.2E-01	5.7E-01	0.00	1458.48
$Aspect$	0.8	-2.0E-04	4.8E-03	-2.8E-02	6.5E-02	6.7E-01	0.00	1458.63
$NDVI{:}S_{\text{veg}}$	0.7	-8.4E-06	1.7E-03	8.7E-02	1.1E-02	1.3E-15	0.15	1410.13
$Wind$	0.7	6.9E-06	5.1E-03	-3.0E-01	6.0E-02	5.7E-07	0.08	1435.84
h_{stream}	0.7	3.9E-04	4.3E-02	-1.0E+00	6.0E-01	7.9E-02	0.01	1455.73
$h_{\text{mixed conifer}}$	0.7	8.1E-05	3.3E-02	6.1E-01	6.2E-01	3.3E-01	0.00	1457.73
$h_{\text{herbaceous}}$	0.7	-5.8E-04	3.7E-02	2.4E-01	6.5E-01	7.1E-01	0.00	1458.66
$h_{\text{graminoid wetland}}$	0.7	2.3E-04	1.5E-02	-6.8E-02	2.4E-01	7.8E-01	0.00	1458.73

Table A.2: Full results of BMA analyses: Shannon's Information Index.

	BMA results			Single-variable regression results				
Coefficient	$p \neq 0$	mean	SD	mean	SE	p	R^2	AIC
Intercept	100	-4.1E+00	1.3E+00	1.4E+00	3.6E-02	<2e-16	0.00	365.99
Day	100	3.5E-02	5.7E-03	3.5E-02	5.8E-03	4.7E-09	0.14	333.29
NDVI.vegS	100	4.9E-02	1.2E-02	4.7E-02	1.1E-02	2.5E-05	0.08	350.06
Hour.To	95	1.3E-02	1.0E-02	2.5E-03	9.2E-04	7.0E-03	0.03	360.64
T_{O}	64.2	-8.9E-02	8.3E-02	2.4E-02	1.1E-02	3.4E-02	0.02	363.47
Hour	58.8	-1.0E-01	1.2E-01	3.4E-02	2.1E-02	1.2E-01	0.01	365.50
$h_{\text{halophytic wetland}}$	46.6	4.8E-01	6.1E-01	9.9E-01	5.4E-01	6.7E-02	0.01	364.61
Sky	19.3	-1.4E-02	3.3E-02	-1.5E-01	3.8E-02	6.3E-05	0.07	351.80
NDVI.E	16.8	-1.2E-04	3.3E-04	4.9E-04	3.5E-04	1.7E-01	0.01	366.08
$h_{\text{white spruce}}$	8.1	1.3E-02	5.3E-02	1.7E-01	1.2E-01	1.5E-01	0.01	365.94
h_{barren}	7.6	5.4E-02	2.3E-01	5.1E-01	5.4E-01	3.5E-01	0.00	367.11
h_{sparse}	5.2	2.4E-02	1.3E-01	-2.0E-01	3.8E-01	6.1E-01	0.00	367.73
$h_{\text{alder-willow}}$	2.6	9.9E-03	8.0E-02	5.0E-01	3.8E-01	1.9E-01	0.01	366.29
$h_{\text{black spruce}}$	2.5	-2.0E-03	1.7E-02	-1.8E-01	8.0E-02	2.2E-02	0.02	362.65
NDVI.Pa	2.3	-5.8E-06	5.2E-05	4.8E-04	2.1E-04	2.3E-02	0.02	362.75
E:T_{A}	1.9	3.6E-06	3.0E-05	9.0E-05	9.1E-05	3.2E-01	0.00	367.00
h_{alder}	1	1.7E-03	2.1E-02	1.5E-01	1.4E-01	3.0E-01	0.00	366.91
T_{A}.Pa	0.6	5.9E-07	9.2E-06	9.3E-05	7.1E-05	1.9E-01	0.01	366.27
NDVI.To	0	0.0E+00	0.0E+00	7.9E-02	2.0E-02	1.2E-04	0.06	353.09
NDVI	0	0.0E+00	0.0E+00	7.8E-01	2.6E-01	2.7E-03	0.04	358.87
T_{A}.To	0	0.0E+00	0.0E+00	6.5E-03	3.2E-03	4.2E-02	0.02	363.80
h_{willow}	0	0.0E+00	0.0E+00	3.8E-01	2.4E-01	1.2E-01	0.01	365.60
E.T_{O}	0	0.0E+00	0.0E+00	1.5E-05	1.0E-05	1.4E-01	0.01	365.84
h_{stream}	0	0.0E+00	0.0E+00	-5.2E-01	3.8E-01	1.8E-01	0.01	366.17
T_{A}	0	0.0E+00	0.0E+00	6.6E-02	5.1E-02	2.0E-01	0.01	366.35
Wind	0	0.0E+00	0.0E+00	7.7E-02	6.2E-02	2.2E-01	0.01	366.45
$h_{\text{mixed conifer}}$	0	0.0E+00	0.0E+00	6.4E-01	5.4E-01	2.4E-01	0.01	366.58

continued on next page...

Table A.2 continued.

Coefficient	$p \neq 0$	mean	SD	mean	SE	p	R^2	AIC
$h_{\text{mixed forest}}$	0	0.0E+00	0.0E+00	8.9E-02	9.1E-02	3.3E-01	0.00	367.02
$Aspect$	0	0.0E+00	0.0E+00	5.5E-02	5.7E-02	3.4E-01	0.00	367.06
$h_{\text{shrub wetland}}$	0	0.0E+00	0.0E+00	1.7E-01	2.1E-01	4.2E-01	0.00	367.33
$E.P_{\text{A}}$	0	0.0E+00	0.0E+00	4.3E-08	6.1E-08	4.9E-01	0.00	367.50
$h_{\text{cottonwood}}$	0	0.0E+00	0.0E+00	2.7E-01	5.4E-01	6.2E-01	0.00	367.74
P_{A}	0	0.0E+00	0.0E+00	4.1E-05	8.8E-05	6.4E-01	0.00	367.77
E	0	0.0E+00	0.0E+00	5.5E-05	1.2E-04	6.5E-01	0.00	367.78
$h_{\text{graminoid wetland}}$	0	0.0E+00	0.0E+00	6.6E-02	1.9E-01	7.4E-01	0.00	367.88
$Slope$	0	0.0E+00	0.0E+00	5.1E-04	2.3E-03	8.3E-01	0.00	367.94
$Area$	0	0.0E+00	0.0E+00	-9.5E-06	6.6E-05	8.9E-01	0.00	367.97
h_{alpine}	0	0.0E+00	0.0E+00	7.1E-03	1.1E-01	9.5E-01	0.00	367.99
$h_{\text{mixed deciduous}}$	0	0.0E+00	0.0E+00	1.6E-02	3.8E-01	9.7E-01	0.00	367.99
$h_{\text{paper birch}}$	0	0.0E+00	0.0E+00	3.2E-03	1.9E-01	9.9E-01	0.00	367.99

Table A.3: Full results of BMA analyses: the Berger-Parker Index. Note that the signs of coefficients are reversed in inverse gamma regressions so that a negative sign indicates a positive relationship.

	BMA results			Single-variable regression results				
Coefficient	$p \neq 0$	mean	SD	mean	SE	p	R^2_{dev}	AIC
Intercept	100	2.0E+00	6.4E-01	4.4E-01	1.4E-02	<2e-16	0.00	668.05
Day	90.9	-8.8E-03	3.5E-03	-1.1E-02	2.3E-03	2.3E-06	0.10	644.28
$NDVI.S_{\text{veg}}$	82.9	-1.1E-02	5.9E-03	-1.4E-02	3.6E-03	1.1E-04	0.06	654.20
$T_{\text{O}}.E$	61.9	-1.6E-05	1.5E-05	-8.3E-06	3.1E-06	7.8E-03	0.03	662.83
$T_{\text{O}}.Hour$	48.8	-6.1E-04	8.3E-04	-9.3E-04	3.3E-04	5.9E-03	0.03	661.39
h_{snow}	33.6	1.8E-01	3.1E-01	5.7E-01	3.5E-01	1.0E-01	0.02	664.77
$Day.E$	28.9	4.8E-07	8.4E-07	-2.2E-07	2.3E-07	3.5E-01	0.00	669.06
E	26.3	7.3E-05	1.4E-04	-3.1E-05	4.0E-05	4.4E-01	0.00	669.36
Sky	25.3	7.0E-03	1.4E-02	5.6E-02	1.5E-02	1.9E-04	0.06	653.72
$NDVI$	17.2	-5.7E-02	1.3E-01	-2.8E-01	9.3E-02	2.7E-03	0.04	658.66
$h_{\text{halophytic wetland}}$	16.3	-4.1E-02	1.0E-01	-1.9E-01	1.2E-01	1.3E-01	0.01	668.19
T_{O}	11.7	1.5E-02	4.8E-02	-9.5E-03	4.1E-03	2.2E-02	0.02	663.84
$Day.T_{\text{O}}$	10.9	-7.6E-05	2.6E-04	-6.9E-05	2.4E-05	4.3E-03	0.04	660.42
$Hour$	8.4	-8.9E-04	6.7E-03	-1.3E-02	8.2E-03	1.0E-01	0.01	666.92
$h_{\text{graminoid wetland}}$	6	-6.1E-03	2.8E-02	-6.1E-02	6.8E-02	3.7E-01	0.00	669.16
$E.P_{\text{A}}$	3	-1.0E-09	6.9E-09	-1.9E-08	1.9E-08	3.1E-01	0.00	668.94
$h_{\text{herbaceous}}$	3	1.2E-02	8.7E-02	3.2E-01	3.7E-01	3.9E-01	0.01	668.80
h_{alpine}	2.8	1.7E-03	1.3E-02	-3.3E-02	3.7E-02	3.8E-01	0.00	669.17
$h_{\text{shrub wetland}}$	2.6	-2.4E-03	1.7E-02	-7.8E-02	6.9E-02	2.6E-01	0.01	668.69
$T_{\text{O}}.P_{\text{A}}$	2.5	-1.1E-07	9.6E-07	-5.0E-06	1.8E-06	6.0E-03	0.03	662.93
h_{sparse}	2.2	-5.2E-03	4.1E-02	6.4E-02	1.8E-01	7.2E-01	0.00	669.88
h_{alder}	1.9	-1.2E-03	1.1E-02	-4.6E-02	5.1E-02	3.7E-01	0.00	669.14
$Wind$	1.5	-3.7E-04	3.9E-03	-1.7E-02	1.8E-02	3.5E-01	0.00	669.10
h_{hemlock}	1.4	1.0E-03	1.2E-02	1.1E-01	7.8E-02	1.6E-01	0.01	667.37
S_{veg}	1.3	-4.2E-05	4.4E-04	-5.8E-03	1.9E-03	2.5E-03	0.04	660.05

continued on next page...

Table A.3 continued.

Coefficient	$p \neq 0$	mean	SD	mean	SE	p	R^2_{dev}	AIC
P_{A}	1.3	1.1E-07	9.0E-06	-2.8E-05	2.9E-05	3.4E-01	0.00	669.01
$T_{\text{O}}.T_{\text{A}}$	1.2	1.9E-05	2.1E-04	-1.3E-03	1.2E-03	2.8E-01	0.01	668.67
$NDVI.P_{\text{A}}$	1	1.3E-06	1.7E-05	-2.0E-04	6.9E-05	3.9E-03	0.04	660.43
$h_{\text{paper birch}}$	1	7.9E-04	1.1E-02	6.4E-02	8.3E-02	4.4E-01	0.00	669.29
h_{barren}	1	-1.4E-03	1.7E-02	-7.3E-02	1.3E-01	5.7E-01	0.00	669.70
$NDVI.T_{\text{O}}$	0.8	-1.7E-04	2.2E-03	-2.6E-02	7.1E-03	2.5E-04	0.06	654.17
h_{willow}	0.6	-3.1E-04	6.0E-03	-1.1E-01	6.9E-02	1.2E-01	0.01	667.68
$Slope$	0.4	-2.0E-06	5.0E-05	-1.1E-03	7.8E-04	1.5E-01	0.01	667.79
$h_{\text{cottonwood}}$	0.4	7.2E-04	1.9E-02	1.4E-01	2.9E-01	6.2E-01	0.00	669.70
$Humidity$	0.4	2.5E-06	7.1E-05	2.8E-04	1.1E-03	8.0E-01	0.00	669.97
$Area$	0.3	4.9E-08	1.4E-06	2.9E-05	2.4E-05	2.4E-01	0.01	668.42
T_{A}	0.3	3.8E-05	1.1E-03	-4.3E-03	2.0E-02	8.2E-01	0.00	669.99
h_{stream}	0.3	-3.2E-04	9.3E-03	8.1E-03	1.6E-01	9.6E-01	0.00	670.05
$NDVI.E$	0.2	-2.1E-07	7.3E-06	-3.4E-04	1.1E-04	2.8E-03	0.04	659.93
$NDVI.T_{\text{A}}$	0.2	-9.6E-06	1.7E-03	-9.1E-02	4.0E-02	2.5E-02	0.02	664.18
$E.T_{\text{A}}$	0.2	-5.4E-08	1.7E-06	-5.1E-05	3.3E-05	1.2E-01	0.01	667.28
$T_{\text{A}}.P_{\text{A}}$	0.2	-5.9E-09	9.3E-07	-3.3E-05	2.8E-05	2.4E-01	0.01	668.40
$h_{\text{alder-willow}}$	0.2	-1.4E-04	5.2E-03	-1.2E-01	1.1E-01	3.0E-01	0.00	668.99
$h_{\text{mixed conifer}}$	0.2	-1.2E-04	6.1E-03	-1.3E-01	1.5E-01	3.7E-01	0.00	669.30
$h_{\text{white spruce}}$	0.2	-3.0E-05	1.7E-03	-2.7E-02	4.3E-02	5.4E-01	0.00	669.61
$h_{\text{mixed forest}}$	0.2	3.7E-05	1.5E-03	-1.4E-02	3.5E-02	6.9E-01	0.00	669.86
$Aspect$	0.2	-1.6E-06	7.3E-04	-7.8E-03	2.2E-02	7.2E-01	0.00	669.90
$h_{\text{mixed deciduous}}$	0.2	7.0E-06	4.8E-03	-3.7E-02	1.4E-01	7.9E-01	0.00	669.97

Appendix B.

Summary of arthropods collected

Table B.1: Frequencies and quantities of arthropod families collected. Note that additional families were sorted (mostly Araneae, Nematocera, and Lepidoptera), but only the families included in analyses are shown.

Class	**Order**	**Family**	f_j	m_j
Gastropoda				20
Arachnida				1,770
	Acarina[1]			387
	Araneae			1,361
	Opiliones			22
		Sclerosomatidae	0.027	22
Myriapoda				2
	Lithobiomorpha			2
		Lithobiidae	0.008	2
Parainsecta				1,066
	Collembola			1,066
		Entomobryidae	0.141	76
		Hypogastruridae	0.004	7
		Isotomidae	0.004	1
		Sminthuridae	0.506	982
Insecta				12,278
	Odonata			3
		Coenagrionidae	0.012	3
	Plecoptera			3
		not sorted		2
		Chloroperlidae	0.004	1
	Orthoptera[2]			13
		Acrididae	0.043	13

Continued on next page...

[1] Approximate. Many of these mites were phoretic on specimens of Diptera (particularly Muscidae) and were difficult to count.

[2] All were immature.

Table B.1 continued.

Class	Order	Family	f_j	m_j
	Psocoptera[3]			64
	Hemiptera			3,011
		not sorted[4]		382
		Acanthosomatidae	0.016	4
		Achilidae	0.004	1
		Anthocoridae	0.008	3
		Aphididae	0.592	1,074
		Cicadellidae	0.494	703
		Delphacidae	0.184	241
		Lygaeidae	0.031	11
		Miridae	0.051	18
		Nabidae	0.004	1
		Psyllidae	0.251	573
	Thysanoptera			91
		not sorted[5]		55
		Phlaeothripidae	0.004	1
		Thripidae	0.086	35
	Neuroptera			28
		not sorted[6]		9
		Chrysopidae	0.012	3
		Coniopterygidae	0.016	6
		Hemerobiidae	0.039	11
	Coleoptera			143
		not sorted[7]		22
		Anobiidae	0.004	1
		Cantharidae	0.125	67
		Carabidae	0.004	1
		Chrysomelidae	0.031	16
		Coccinellidae	0.012	3
		Curculionidae[8]	0.039	12
		Elateridae	0.016	5
		Lathridiidae	0.008	2

Continued on next page...

[3] All were immature.
[4] Immature specimens.
[5] Immature specimens.
[6] Immature specimens.
[7] Immature specimens.
[8] Includes Scolytinae.

Table B.1 continued.

Class	Order	Family	f_j	m_j
		Leiodidae	0.004	1
		Lycidae	0.004	1
		Pythidae	0.004	1
		Scarabaeidae	0.004	1
		Scirtidae	0.004	1
		Staphylinidae	0.027	9
	Diptera			8,074
		not sorted[9]		2,530
		Agromyzidae	0.082	55
		Anisopodidae	0.016	4
		Anthomyiidae	0.169	95
		Anthomyzidae	0.047	21
		Asteiidae	0.004	4
		Bibionidae	0.114	85
		Chamaemyiidae	0.020	10
		Chloropidae	0.004	52
		Clusiidae	0.004	1
		Culicidae	0.761	3,697
		Dolichopodidae	0.024	76
		Drosophilidae	0.047	21
		Dryomyzidae	0.004	1
		Empididae[10]	0.345	212
		Ephydridae	0.012	135
		Heleomyzidae	0.047	20
		Lauxaniidae	0.161	119
		Micropezidae	0.008	3
		Muscidae[11]	0.314	516
		Otitidae	0.008	17
		Phoridae	0.133	87
		Pipunculidae	0.035	18
		Psilidae	0.031	22
		Rhagionidae	0.008	2
		Scathophagidae	0.082	33
		Sciomyzidae	0.075	29
		Sepsidae	0.004	1
		Simuliidae	0.224	134
		Sphaeroceridae	0.047	27
		Stratiomyidae	0.012	3

Continued on next page...

[9] Damaged material and several families of Nematocera.
[10] Includes Hybotinae.
[11] Includes Fanninae.

Table B.1 continued.

Class	**Order**	**Family**	f_j	m_j
		Syrphidae	0.075	33
		Tabanidae	0.008	2
		Tachinidae	0.024	7
		Tephritidae	0.004	2
	Trichoptera			4
	Lepidoptera			168
	Hymenoptera			675
		not sorted[12]		21
		immature Symphyta		150
		Aphelinidae	0.004	1
		Apidae	0.024	6
		Argidae	0.004	2
		Bethylidae	0.004	1
		Braconidae	0.314	128
		Ceraphronidae	0.008	2
		Diapriidae	0.106	38
		Dryinidae	0.031	9
		Encyrtidae	0.024	7
		Eulophidae	0.063	29
		Eurytomidae	0.016	4
		Formicidae	0.063	52
		Ichneumonidae	0.345	181
		Platygasteridae	0.024	8
		Pteromalidae	0.047	12
		Scelionidae	0.008	2
		Tenthredinidae	0.059	17
		Torymidae	0.020	5
Total				15,136
Analysis Total				9,961

[12]Damaged material.

Appendix C.

Tabular results of simulations

Table C.1: Failure rate of the rotating panel design for a single season. Values are the percentage of simulations in which PRESENCE failed. Failures were defined as fitting of either Ψ or p with a value of 1 or 0.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Ψ	0.1	98	85	71	50	27	17	13	24	40
	0.2	85	62	38	16	10	3	2	2	13
	0.3	86	59	22	10	1	0	0	1	2
	0.4	88	48	15	13	1	0	0	0	1
	0.5	76	41	14	10	2	0	0	0	1
	0.6	68	42	25	12	4	1	0	0	0
	0.7	67	31	32	18	13	2	0	0	1
	0.8	71	34	30	27	16	5	2	0	0
	0.9	69	47	40	35	32	20	14	1	0

Table C.2: Failure rate of the periodic census design for a single season. Values are the percentage of simulations in which PRESENCE failed. Failures were defined as fitting of either Ψ or p with a value of 1 or 0.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Ψ	0.1	65	39	13	0	0	0	0	0	2
	0.2	60	20	0	0	0	0	0	0	0
	0.3	44	8	0	0	0	0	0	0	0
	0.4	39	13	1	0	0	0	0	0	0
	0.5	22	8	3	0	0	0	0	0	0
	0.6	26	17	2	0	0	0	0	0	0
	0.7	37	22	7	0	0	0	0	0	0
	0.8	36	19	14	3	1	1	0	0	0
	0.9	52	36	27	14	6	1	0	0	0

Table C.3: Bias of estimates of Ψ for the rotating panel design for a single season, naïve estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	-0.081	-0.063	-0.047	-0.035	-0.022	-0.021	-0.006	-0.007	-0.003
	0.2	-0.164	-0.125	-0.107	-0.068	-0.054	-0.037	-0.019	-0.010	-0.006
	0.3	-0.246	-0.192	-0.150	-0.117	-0.059	-0.036	-0.030	-0.013	0.002
	0.4	-0.326	-0.254	-0.198	-0.151	-0.098	-0.063	-0.035	-0.019	0.003
Ψ	0.5	-0.407	-0.325	-0.246	-0.172	-0.131	-0.082	-0.058	-0.022	-0.008
	0.6	-0.484	-0.392	-0.294	-0.214	-0.142	-0.089	-0.054	-0.035	-0.006
	0.7	-0.564	-0.451	-0.340	-0.254	-0.178	-0.108	-0.065	-0.029	-0.005
	0.8	-0.649	-0.513	-0.395	-0.285	-0.215	-0.117	-0.069	-0.027	-0.013
	0.9	-0.725	-0.578	-0.445	-0.326	-0.215	-0.147	-0.072	-0.038	-0.010

Table C.4: Bias of estimates of Ψ for the periodic census design for a single season, naïve estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	-0.081	-0.063	-0.049	-0.037	-0.022	-0.017	-0.009	-0.004	-0.001
	0.2	-0.163	-0.131	-0.100	-0.076	-0.052	-0.035	-0.017	-0.008	-0.003
	0.3	-0.243	-0.192	-0.146	-0.104	-0.079	-0.049	-0.026	-0.013	-0.002
	0.4	-0.321	-0.252	-0.198	-0.142	-0.101	-0.063	-0.033	-0.013	-0.005
Ψ	0.5	-0.403	-0.320	-0.242	-0.180	-0.132	-0.078	-0.046	-0.024	-0.015
	0.6	-0.486	-0.383	-0.296	-0.219	-0.151	-0.097	-0.052	-0.022	-0.003
	0.7	-0.564	-0.446	-0.346	-0.253	-0.175	-0.108	-0.065	-0.031	-0.002
	0.8	-0.646	-0.512	-0.395	-0.294	-0.202	-0.133	-0.071	-0.030	-0.006
	0.9	-0.728	-0.580	-0.438	-0.322	-0.228	-0.143	-0.083	-0.037	-0.009

Table C.5: Bias of estimates of Ψ for the rotating panel design for a single season, PRESENCE estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	-0.056	-0.022	0.015	0.021	0.022	0.008	0.017	0.012	0.018
	0.2	-0.113	-0.044	-0.036	0.027	0.029	0.015	0.020	0.008	0.007
	0.3	-0.209	-0.051	0.009	0.028	0.058	0.053	0.006	0.004	0.009
	0.4	-0.202	-0.058	0.012	0.025	0.029	0.017	0.012	0.000	0.008
Ψ	0.5	-0.264	-0.087	0.037	0.020	0.002	0.009	-0.001	0.002	0.001
	0.6	-0.330	-0.079	-0.004	-0.006	0.040	0.036	0.007	-0.008	0.001
	0.7	-0.327	-0.132	-0.016	-0.008	0.004	0.024	0.011	0.002	0.003
	0.8	-0.366	-0.199	-0.093	-0.034	-0.009	0.013	0.008	0.010	-0.004
	0.9	-0.305	-0.221	-0.144	-0.075	-0.028	-0.023	0.002	0.003	-0.011

Table C.6: Bias of estimates of Ψ for the periodic census design for a single season, PRESENCE estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	0.421	0.021	0.007	0.020	0.011	0.005	-0.001	0.004	0.000
	0.2	-0.069	0.023	0.031	0.013	0.011	0.004	-0.005	0.002	-0.001
	0.3	-0.084	0.034	0.033	-0.002	0.007	0.006	-0.002	0.004	0.003
	0.4	-0.023	0.026	0.055	0.018	-0.002	-0.001	-0.001	-0.001	-0.006
Ψ	0.5	-0.038	0.027	0.006	0.012	0.003	0.007	-0.001	0.001	-0.002
	0.6	-0.088	0.000	0.012	0.018	-0.001	0.004	0.006	0.000	-0.002
	0.7	-0.137	-0.022	0.004	0.012	0.004	0.009	-0.001	-0.003	-0.005
	0.8	-0.166	-0.050	-0.032	-0.006	0.000	0.007	0.002	0.005	0.004
	0.9	-0.221	-0.071	-0.060	-0.023	0.007	-0.002	0.001	0.000	0.000

Table C.7: Standard deviation of estimates of Ψ for the rotating panel design for a single season, naïve estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	0.018	0.028	0.032	0.032	0.036	0.039	0.037	0.041	0.040
	0.2	0.026	0.040	0.038	0.049	0.052	0.050	0.052	0.052	0.055
	0.3	0.029	0.042	0.049	0.054	0.067	0.056	0.062	0.066	0.068
	0.4	0.034	0.049	0.060	0.065	0.071	0.078	0.074	0.064	0.060
Ψ	0.5	0.037	0.047	0.058	0.066	0.066	0.072	0.070	0.070	0.075
	0.6	0.045	0.054	0.063	0.067	0.070	0.076	0.072	0.062	0.063
	0.7	0.050	0.056	0.066	0.069	0.070	0.069	0.066	0.069	0.060
	0.8	0.051	0.061	0.070	0.066	0.078	0.071	0.056	0.062	0.057
	0.9	0.056	0.061	0.078	0.069	0.067	0.065	0.052	0.047	0.038

Table C.8: Standard deviation of estimates of Ψ for the periodic census design for a single season, naïve estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	0.010	0.010	0.015	0.015	0.017	0.018	0.019	0.017	0.019
	0.2	0.012	0.015	0.019	0.020	0.020	0.025	0.025	0.023	0.024
	0.3	0.014	0.019	0.025	0.024	0.025	0.033	0.027	0.027	0.028
	0.4	0.017	0.022	0.026	0.028	0.029	0.030	0.030	0.032	0.026
Ψ	0.5	0.020	0.023	0.025	0.029	0.029	0.032	0.033	0.030	0.031
	0.6	0.023	0.023	0.029	0.030	0.033	0.030	0.027	0.031	0.031
	0.7	0.022	0.032	0.029	0.028	0.034	0.033	0.029	0.025	0.028
	0.8	0.020	0.029	0.032	0.031	0.031	0.028	0.026	0.026	0.024
	0.9	0.026	0.028	0.026	0.033	0.026	0.027	0.024	0.022	0.019

Table C.9: Standard deviation of estimates of Ψ for the rotating panel design for a single season, PRESENCE estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	0.000	0.044	0.075	0.051	0.050	0.046	0.044	0.037	0.037
	0.2	0.047	0.082	0.079	0.118	0.093	0.089	0.082	0.057	0.052
	0.3	0.058	0.171	0.165	0.163	0.135	0.118	0.076	0.068	0.067
	0.4	0.128	0.161	0.193	0.147	0.135	0.119	0.084	0.069	0.061
Ψ	0.5	0.140	0.180	0.189	0.146	0.108	0.107	0.088	0.075	0.075
	0.6	0.162	0.233	0.184	0.146	0.126	0.129	0.082	0.068	0.062
	0.7	0.226	0.200	0.178	0.148	0.127	0.110	0.081	0.076	0.061
	0.8	0.202	0.186	0.169	0.137	0.123	0.094	0.072	0.066	0.058
	0.9	0.219	0.175	0.148	0.115	0.076	0.076	0.058	0.055	0.077

Table C.10: Standard deviation of estimates of Ψ for the periodic census design for a single season, PRESENCE estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	0.420	0.150	0.064	0.056	0.048	0.027	0.021	0.018	0.017
	0.2	0.136	0.124	0.120	0.059	0.042	0.030	0.029	0.025	0.025
	0.3	0.127	0.150	0.114	0.065	0.045	0.038	0.030	0.031	0.031
	0.4	0.200	0.175	0.152	0.079	0.045	0.039	0.035	0.036	0.032
Ψ	0.5	0.192	0.181	0.095	0.089	0.059	0.045	0.038	0.031	0.029
	0.6	0.206	0.149	0.110	0.086	0.063	0.043	0.040	0.029	0.030
	0.7	0.194	0.146	0.121	0.080	0.064	0.054	0.035	0.030	0.028
	0.8	0.183	0.131	0.107	0.077	0.065	0.057	0.041	0.035	0.039
	0.9	0.171	0.099	0.078	0.071	0.048	0.048	0.031	0.028	0.018

Table C.11: Failure rate of the rotating panel design over four years. Values are the percentage of simulations in which PRESENCE failed. Failures were defined as fitting of either Ψ or p with a value of 1 or 0.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Ψ	0.1	82	68	39	21	6	2	2	7	17
	0.2	78	53	18	4	1	0	0	0	5
	0.3	79	31	10	0	0	0	0	0	0
	0.4	67	26	9	5	0	0	0	0	0
	0.5	61	30	5	4	1	0	0	0	0
	0.6	53	27	15	6	1	0	0	0	1
	0.7	52	33	18	7	1	1	0	0	2
	0.8	58	49	36	17	5	2	0	0	7
	0.9	50	45	43	46	31	12	8	4	12

Table C.12: Bias of estimates of ε for the rotating panel design over four years, PRESENCE estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Ψ	0.1	0.670	0.440	0.331	0.315	0.369	0.424	0.473	0.554	0.669
	0.2	0.494	0.370	0.316	0.426	0.461	0.512	0.588	0.683	0.772
	0.3	0.424	0.353	0.407	0.450	0.513	0.559	0.593	0.624	0.645
	0.4	0.384	0.400	0.415	0.456	0.473	0.518	0.540	0.581	0.576
	0.5	0.359	0.357	0.395	0.425	0.393	0.412	0.446	0.435	0.457
	0.6	0.395	0.335	0.293	0.330	0.329	0.333	0.343	0.339	0.409
	0.7	0.367	0.302	0.244	0.238	0.258	0.247	0.258	0.246	0.306
	0.8	0.394	0.284	0.223	0.166	0.135	0.173	0.193	0.176	0.172
	0.9	0.323	0.220	0.159	0.091	0.116	0.097	0.085	0.096	0.113

Table C.13: Standard Deviation of estimates of ε for the rotating panel design over four years, PRESENCE estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Ψ	0.1	0.303	0.359	0.288	0.246	0.207	0.165	0.176	0.197	0.222
	0.2	0.344	0.269	0.174	0.153	0.113	0.121	0.100	0.118	0.131
	0.3	0.301	0.207	0.158	0.118	0.109	0.089	0.088	0.116	0.129
	0.4	0.281	0.186	0.129	0.121	0.096	0.090	0.081	0.091	0.152
	0.5	0.242	0.163	0.169	0.112	0.135	0.121	0.090	0.082	0.099
	0.6	0.214	0.188	0.161	0.143	0.115	0.104	0.086	0.084	0.100
	0.7	0.219	0.187	0.167	0.137	0.121	0.095	0.090	0.108	0.151
	0.8	0.187	0.172	0.156	0.147	0.118	0.102	0.083	0.086	0.126
	0.9	0.185	0.144	0.144	0.121	0.099	0.103	0.082	0.078	0.072

Table C.14: Bias of estimates of γ for the rotating panel design over four years, PRESENCE estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Ψ	0.1	0.006	0.019	0.046	0.069	0.049	0.062	0.051	0.064	0.095
	0.2	0.024	0.105	0.142	0.146	0.147	0.149	0.145	0.190	0.193
	0.3	0.070	0.217	0.210	0.243	0.226	0.225	0.268	0.270	0.278
	0.4	0.128	0.219	0.301	0.316	0.342	0.333	0.350	0.383	0.399
	0.5	0.205	0.35	0.378	0.392	0.45	0.447	0.431	0.445	0.453
	0.6	0.201	0.395	0.465	0.515	0.500	0.516	0.529	0.547	0.640
	0.7	0.278	0.415	0.534	0.544	0.539	0.613	0.640	0.608	0.710
	0.8	0.245	0.442	0.545	0.500	0.578	0.699	0.750	0.681	0.731
	0.9	0.317	0.492	0.561	0.522	0.664	0.765	0.650	0.826	0.912

Table C.15: Standard deviation of estimates of γ for the rotating panel design over four years, PRESENCE estimates.

		p								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Ψ	0.1	0.032	0.080	0.071	0.090	0.074	0.060	0.066	0.069	0.070
	0.2	0.057	0.162	0.141	0.140	0.099	0.103	0.071	0.063	0.056
	0.3	0.078	0.171	0.179	0.138	0.102	0.077	0.069	0.062	0.075
	0.4	0.187	0.162	0.174	0.122	0.103	0.092	0.081	0.074	0.093
	0.5	0.177	0.199	0.176	0.113	0.124	0.091	0.064	0.065	0.099
	0.6	0.141	0.214	0.138	0.102	0.090	0.094	0.095	0.104	0.16
	0.7	0.211	0.204	0.137	0.133	0.109	0.129	0.108	0.112	0.216
	0.8	0.159	0.191	0.132	0.152	0.174	0.125	0.124	0.152	0.292
	0.9	0.222	0.160	0.148	0.188	0.131	0.132	0.270	0.163	0.136

Appendix D.

R code used for running occupancy simulations using program PRESENCE

The R code I wrote for generating simulated occupancy data and fitting models using program PRESENCE is included below. There are two scripts: a single-season script and a multi-season script. These scripts were saved as text files and loaded into R using R's `source()` function. Percent symbols (%) mark line breaks imposed to fit the script within the margins of the page.

Single-season simulations

```
### Functions to run simulations through program PRESENCE.
   ### Matthew Bowser
   ### 27-VIII-2008

   ### Function to generate simulated data.
   generate.data <- function(n,  K,  psi,  p,  nsims)
    {
    require(Rlab)
    l.psi <- length(psi)
    l.p <- length(p)
    l.K <- length(K)
    data <- array(data = NA,  dim = c(n,  K,  l.psi,  l.p,  l.K,  nsims))
    for (i.psi in 1:l.psi)
     {
     for (i.p in 1:l.p)
      {
      for (i.K in 1:l.K)
       {
       for (i.nsims in 1:nsims)
        {
        zz <- rbern(n,  psi[i.psi])
        for (i.i.K in 1:K[i.K])
         {
         bb <- rbern(n,  p[i.p])
         data[, i.i.K,  i.psi,  i.p,  i.K,  i.nsims] <- zz*bb
         }
        }
       }
      }
     }
    data
    }

   ### Functions to run PRESENCE,  fitting occupancy models to the data.
```

```
### Function to make .pao data files.
make.pao <- function(data,  data.file) ### Data has to be a two-dimensional %
   array with dimensions n and K.
 {
 n <- length(data[, 1])
 K <- length(data[1, ])
 line1 <- c(n,  K)
 write(line1,  data.file,  ncolumns=2,  append=FALSE,  sep="\t")
 write(t(data),  data.file,  ncolumns=2,  append=TRUE,  sep="\t")
 pars <- c(0,  0,  K)
 write(pars,  data.file,  ncolumns=1,  append=TRUE,  sep="\t")
 write("TITLE:test",  data.file,  append=TRUE)
 site <- rep("site",  n)
 num <- 1:n
 sites <- cbind(site,  num)
 write(t(sites),  data.file,  ncolumns=2,  append=TRUE,  sep=" ")
 seasons <- rep(1,  K)
 visits <- 1:K
 surveys <- cbind(seasons,  visits)
 write(t(surveys),  data.file,  ncolumns=2,  append=TRUE,  sep="-")
 ### I think it is a data checksum,  but I do not know how to generate it.
 write(1514,  data.file,  append=TRUE)
 }

### Funciton to run 100 simulations for a given set of (scalar) values of n,  %
   K,  psi,  and p.
run.PRESENCE <- function(data,  working.directory,  PRESENCE.directory,  %
   file.root) ### Data has to be a three-dimentional array with dimensions n, %
    K,  and nsims.
 {
 ### File names.
 data.file <- paste(file.root,  ".pao",  sep="")
 dm.file <- paste(file.root,  ".dm",  sep="")
 out.file <- paste(file.root,  ".pa2.out",  sep="")
 nsims <- length(data[1, 1, ])
 ### Making an array to hold the results.
 results <- array(NA,  dim=c(2,  nsims))
 ### Going for it.
 for (i.nsims in 1:nsims)
  {
  ### Make a data file.
  setwd(working.directory)
  make.pao(data[, , i.nsims],  data.file)
  ### Making the design matrix file.
  setwd(working.directory)
  write("",  dm.file)
  ### Run PRESENCE.
  setwd(PRESENCE.directory)
  n <- length(data[, 1, 1])
  K <- length(data[1, , 1])
  command <- paste("presence.exe i=",  working.directory,  "/",  data.file,  %
   " l=",  working.directory,  "/",  out.file,  " name=\"simple model\" %
```

```
   model=11c lmt=100 N=",  n,  ",  T=",  K,  ",  NPar=2,  TSpecific=0",  %
   sep="")
  system(command = command,  wait=TRUE)
  setwd(working.directory)
  output <- scan(file=out.file,  what=character())
  cc <- match("(Psi)",  output)
  dd <- match("p",  output)
  results[1, i.nsims] <- as.numeric(output[cc + 2])
  results[2, i.nsims] <- as.numeric(output[dd + 8])
  }
 results
 }

### Function that takes care of iterating through all the values of n,  K,  %
   psi,  and p and fitting the models to the simulated data.
run.PRESENCE.sims <- function(data,  nsims,  working.directory,  file.root,  %
   PRESENCE.directory)
 {
 l.psi <- length(data[1, 1, , 1, 1, 1])
 l.p <- length(data[1, 1, 1, , 1, 1])
 l.K <- length(data[1, 1, 1, 1, , 1])
 ### Making an array to hold the results.
 results <- array(data = NA,  dim = c(l.psi,  l.p,  l.K,  2,  nsims))
 ### Running PRESENCE.
 for(i.psi in 1:l.psi)
  {
  for (i.p in 1:l.p)
   {
   for (i.K in 1:l.K)
    {
    results[i.psi,  i.p,  i.K, , ] <-
    run.PRESENCE(data = data[ ,  ,  i.psi,  i.p,  i.K,  ],  %
   working.directory,  PRESENCE.directory,  file.root)
    }
   }
  }
 results
 }

### Master function that calls everything else.
run.sims.array <- function(n,  K,  psi,  p,  nsims,  working.directory,  %
   file.root,  PRESENCE.directory)
 {
 t <- proc.time()[3]
 ### Go to the working directory.
 setwd(working.directory)
 ### Generate data and save it.
 data <- generate.data(n,  K,  psi,  p,  nsims)
 save(data,  file=paste(file.root,  "data.roj",  sep=""))
 ### Run the simulations and save the results.
 results <- run.PRESENCE.sims(data,  nsims,  working.directory,  file.root,  %
   PRESENCE.directory)
```

```
    save(results,  file="results")
    s <- proc.time()[3] - t
    gg <- length(K) * length(psi) * length(p) * nsims
    kk <- n*K
    print(paste("It took",  round(s, 1),  "seconds to run",  gg,  "simulations,  %
      each simulation including",  kk,  "observations."))
    print(paste("That means it took about",  round(s/gg, 3),  "seconds per %
      simulation."))
    setwd(working.directory)
    }
```

Multi-season simulations

```
### Function to generate data.
   generate.multiseason.data <- function(n,  K,  seasons,  psi,  p,  epsilon,  %
     gamma,  panel)
    {
    ### Make sure the Rlab package is loaded.
    require(Rlab)
    ### Generating initial occupancy states.
    occupancy <- array(data=NA,  dim=c(n, seasons))
    occupancy[, 1] <- rbern(n,  psi)
    for (i.season in 2:seasons)
     {
     for (i.n in 1:n)
      {
      ### Colonization
      if (occupancy[i.n, i.season-1] == 0)
       {
       occupancy[i.n, i.season] <- rbern(1, gamma)
       }
      ### Extinction
      if (occupancy[i.n, i.season-1] == 1)
       {
       occupancy[i.n, i.season] <- 1 - rbern(1, epsilon)
       }
      }
     }

    ### Generating potential observations.
    observations <- array(data=NA,  dim=c(n, K, seasons))
    for (i.seasons in 1:seasons)
     {
     ps <- array(data=rbern(n*K,  p),  dim=c(n, K))
     observations[, , i.seasons] <- ps * occupancy[, i.seasons]
     }

    ### Making a panel structure of the right size.
    use.panel <- array(data=panel,  dim=c(n, seasons))

    use.observations <- observations
```

```
 for (i.seasons in 1:seasons)
  {
   for (i.n in 1:n)
   {
   ### Want to refer to a location
   if (use.panel[i.n, i.seasons] == 1)
    {
    use.observations[i.n, , i.seasons] <- observations[i.n, , i.seasons]
    }
   else
    {
    use.observations[i.n, , i.seasons] <- "."
    }
   }
  }

 ### Output.
 use.observations
 }


### Function to make .pao data files.
make.pao.multi <- function(data.file,  data,  n,  K,  seasons)
 {
 data.eh <- array(data=data,  dim=c(n,  K*seasons))
 line1 <- c(n,  K*seasons)
 write(line1,  data.file,  ncolumns=2,  append=FALSE,  sep="\t")
 write(t(data.eh),  data.file,  ncolumns=K*seasons,  append=TRUE,  sep="\t")
 pars <- c(0,  0,  K)
 write(pars,  data.file,  ncolumns=1,  append=TRUE,  sep="\t")
 write("test",  data.file,  append=TRUE)
 site <- rep("site",  n)
 num <- 1:n
 sites <- cbind(site,  num)
 write(t(sites),  data.file,  ncolumns=2,  append=TRUE,  sep=" ")
 seasons.list <- 1:seasons
 K.list <- 1:K
 surveys <- cbind(expand.grid(K.list, seasons.list)[, 2], expand.grid(K.list, %
   seasons.list)[, 1])
 write(t(surveys),  data.file,  ncolumns=2,  append=TRUE,  sep="-")
 ### I think it is a data checksum,  but I do not know how to generate it.
 write(8000,  data.file,  append=TRUE)
 }

### Function to make .dm files.
make.dm.multi.constants <- function(dm.file,  K,  seasons)
 {
 list.pars <- 0:5
 l <- rep(NA,  4)
 l[1] <- 2
 l[2] <- seasons
 l[3] <- l[2]
```

```
 l[4] <- K*seasons + 1
 l[5] <- 0
 l[6] <- l[5]
 e <- c(rep(2, 4),  rep(0, 2))
 dimensions <- cbind(list.pars,  l,  e)
 seasons.list <- 1:seasons
 K.list <- 1:K
 surveys <- cbind(expand.grid(K.list, seasons.list)[, 2], expand.grid(K.list, %
   seasons.list)[, 1])
 ss <- rep(NA,  K*seasons)
 for (i.surveys in 1:(K*seasons))
  {
  ss[i.surveys] <- paste(surveys[i.surveys, 1],  surveys[i.surveys, 2],  %
   sep="-")
  }
 write(dimensions[1, ],  dm.file,  ncolumns=3,  append=FALSE,  sep=" ")
 write("-, a1,
psi1, 1, ",  dm.file,  append=TRUE)
 write(dimensions[2, ],  dm.file,  ncolumns=3,  append=TRUE,  sep=" ")
 write("-, b1, ",  dm.file,  append=TRUE)
 for (i.seasons in 1:(seasons-1))
  {
  line <- paste("gam",  i.seasons,  ", 1, ",  sep="")
  write(line,  dm.file,  append=TRUE)
  }
 write(dimensions[3, ],  dm.file,  ncolumns=3,  append=TRUE,  sep=" ")
 write("-, c1, ",  dm.file,  append=TRUE)
 for (i.seasons in 1:(seasons-1))
  {
  line <- paste("eps",  i.seasons,  ", 1, ",  sep="")
  write(line,  dm.file,  append=TRUE)
  }
 write(dimensions[4, ],  dm.file,  ncolumns=3,  append=TRUE,  sep=" ")
 write("-, d1, ",  dm.file,  append=TRUE)
 for (i.ss in 1:(K*seasons))
  {
  line <- paste("P[",  ss[i.ss],  "], 1, ",  sep="")
  write(line,  dm.file,  append=TRUE)
  }
 write(t(dimensions[5:6, ]),  dm.file,  ncolumns=3,  append=TRUE,  sep=" ")
 }

### Now need a function to run PRESENCE.
call.presence <- function(PRESENCE.directory,  working.directory,  data.file, %
    dm.file,  out.file,  n,  K)
 {
 command <- paste("presence.exe i=",  working.directory,  "/",  data.file,  " %
   l=",  working.directory,  "/",  out.file,  " name=\"simple model\" %
   model=2000",  " j=",  working.directory,  "/",  dm.file,  " lmt=200",  %
   sep="")
 setwd(PRESENCE.directory)
 system(command = command,  wait=TRUE)
```

```
 setwd(working.directory)
 command
 }


### Function to transform untransformed estimates back to their original scale.
untrans <- function(aa)
 {
 bb <- exp(aa)/(exp(aa) + 1)
 bb
 }
### Function to extract results.
extract.results <- function(out.file)
 {
 output <- scan(file=out.file,  what=character())
 A1 <- match("A1",  output)
 psi <- untrans(as.numeric(output[A1 + 3]))
 B1 <- match("B1",  output)
 gamma <- untrans(as.numeric(output[B1 + 3]))
 C1 <- match("C1",  output)
 epsilon <- untrans(as.numeric(output[C1 + 4]))
 D1 <- match("D1",  output)
 p <- untrans(as.numeric(output[D1 + 3]))
 results <- array(data=c(psi,  p,  epsilon,  gamma),  dim=4)
 results
 }


### Now need master function to call all the others.
run.multi.sims <- function(n,  K,  seasons,  psi,  p,  epsilon,  gamma,  %
   panel,  nsims,  file.root,  PRESENCE.directory,  working.directory)
 {
 setwd(working.directory)
 ### First need to make something to hold all of the data.
 multi.sims <- list(n=n,  K=K,  seasons=seasons,  psi=psi,  p=p,  %
   epsilon=epsilon,  gamma=gamma,  panel=panel,  nsims=nsims)
 multi.sims$data <- array(data=NA,  dim=c(n, K*seasons, nsims))
 ### Make the .dm file.
 dm.file <- paste(file.root,  ".dm",  sep="")
 make.dm.multi.constants(dm.file=dm.file,  K=K,  seasons=seasons)
 ### Make an array to hold the results.
 multi.sims$results <- array(data=NA,  dim=c(4, nsims))
 ### Go for it!
 for (sim in 1:nsims)
  {
  ### Generating data.
  multi.sims$data[, , sim] <- generate.multiseason.data(n=n,  K=K,  %
   seasons=seasons,  psi=psi,  p=p,  epsilon=epsilon,  gamma=gamma,  %
   panel=panel)
  ### Making .pao file.
  data.file <- paste(file.root,  sim,  ".pao",  sep="")
```

```
  make.pao.multi(data.file=data.file,  data=multi.sims$data[, , sim],  n=n,  %
   K=K,  seasons=seasons)
  ### Running PRESENCE.
  out.file = paste(file.root,  sim,  ".out",  sep="")
  call.presence(PRESENCE.directory=PRESENCE.directory,  %
   working.directory=working.directory,  data.file=data.file,  %
   dm.file=dm.file,  out.file=out.file,  n=n,  K=K)
  ### Extracting and storing the results.
  multi.sims$results[, sim] <- extract.results(out.file=out.file)
  }
 multi.sims
 }


### Example of use.
#panel <- array(data=c(1, 1, 0, 0, 0, 0, 1, 1),  dim=c(4, 2))
#test <- run.multi.sims(n=4,  K=2,  seasons=6,  psi=0.8,  p=0.8,  %
   epsilon=0.1,  gamma=0.2,  panel=panel,  nsims=10,  #file.root="test01",
#PRESENCE.directory="C:/Program Files/PRESENCE",  %
   working.directory="C:/tmp/20081006multiseason")
```